CATÓLICA

UNIVERSIDADE CATÓLICA PORTUGUESA | PORTO

Escola Superior de Biotecnologia

Measurement Uncertainty in Screening Immunoassays in Blood Establishments

Thesis submitted to the *Universidade Católica Portuguesa* to attain the degree of PhD in Biotechnology with specialization in Microbiology

Paulo António Rodrigues Pereira

December 2015

This page intentionally left blank

This thesis is dedicated to my Grandparents,

António Rodrigues (b. 1921, d. 2007) and Maria Madalena dos Santos Rodrigues (b. 1931, d. 2014).

All that happened in my life was possible only because of their education, love and belief.

This page intentionally left blank

RESUMO

No contexto dos serviços de sangue, a segurança pós-transfusão do recetor de sangue é considerada um tópico maior. Considerando a segurança dos recetores de sangue, estes serviços deverão testar um conjunto de agentes transmissíveis com prevalência significante na população. O papel dos resultados dos testes de rastreio na validação das dádivas de sangue é crítico para a garantia da segurança pós-transfusão. Nos países da União Europeia, o rastreio é requerido pela diretiva 2001/83/EC e nos EUA pelas "normas para bancos de sangue e serviços de transfusão".

Dada a incerteza associada aos testes de rastreio, um resultado reportado poderá ser falso. Se um resultado falso positivo não tem consequência para a segurança pós-transfusão, já um resultado falso negativo tem um grande impacto já que existe uma probabilidade elevada de o recetor do componente sanguíneo ser infetado. Consequentemente, os laboratórios de rastreio devem considerar a avaliação da incerteza de medição nos esquemas de controlo de qualidade. Contudo, os requisitos da União Europeia ou dos EUA relacionados com incerteza são genéricos.

Esta tese reúne um conjunto de artigos com discussões acerca de métodos para a determinação da incerteza da medição em imunoensaios de rastreio, assim como de abordagens complementares e modelos de exatidão diagnóstica. A utilização do Guia para a Expressão da Incerteza na Medição (GUM) não é sistemática nos laboratórios de rastreio dos serviços de sangue, tal como noutros laboratórios clínicos. Contudo, têm sido usados nos serviços de sangue outras estimativas de incerteza usando métodos estatísticos. Assim, a discussão reúne, não só modelos de acordo com os princípios do GUM focados na incerteza de resultados próximos ou iguais ao "cutoff" (ponto de decisão clínica), mas também modelos complementares: erro total analítico, período de janela seronegativo e modelo delta para uma estimativa inicial do impacto de resultados incertos no orçamento de um serviço de sangue. Na perspetiva da incerteza diagnóstica reúnem-se os modelos de exatidão diagnóstica para a determinação da sensibilidade clínica, especificidade clínica e área sob a curva da característica de operação do recetor. Complementariamente, são discutidos modelos para a estimação da concordância de resultados entre testes.

Para ilustrar os vários métodos de cálculo foram usados resultados de um imunoensaio quimiluminescente comum para antivírus da hepatite C. Os cálculos mostram incerteza expandida em torno do "cutoff" de 21 a 36 %, tendo sido usado para a estimativa da zona de rejeição [0.70, +∞[, na qual se registaram somente 0.19 % dos resultados de 9805 amostras. O erro total analítico foi de 23 %. O modelo de período de janela estimou um período seronegativo de 97 dias, considerando a zona de rejeição do imunoensaio. O valor delta indicou que a amostragem de doentes testada no "teste 1" tem uma menor probabilidade para gerar resultados indeterminados, não tendo sido evidenciada uma diferença estatisticamente significativa do "teste 1 e "teste 2" gerarem resultados indeterminados na amostra de não doentes. Os intervalos de confiança a 95 % para a sensibilidade e especificidade clínicas foram, respetivamente, de $IC_{S,95\%}$ = [88.3, 100 %] e $IC_{E,95\%}$ = [98.5, 100 %], tendo-se obtido uma área sob a curva da característica de operação do recetor entre 0.99 e 1.00. A concordância total entre os resultados de dois imunoensaios para antivírus da hepatite C foi de 98 a 99 %, tendo sido a concordância entre resultados positivos de 87.9 a 100 % e a concordância entre resultados negativos de 97.8 a 99.9 %. A aplicação do intervalo de confiança a 95 % às estimativas de probabilidades e concordâncias é similar ao conceito de incerteza expandida da incerteza da medição.

Todos os métodos apresentados têm aplicação na avaliação da incerteza de medição e incerteza diagnóstica de um imunoensaio, embora tenham condições de utilização diferentes. Recomenda-se um esquema para a seleção dos modelos para a estimativa da incerteza em laboratórios de rastreio de serviços de sangue, atendendo ao papel de cada modelo na segurança pós-transfusão.

Palavras-chave: erro total analítico, exatidão diagnóstica, GUM, imunoensaio de rastreio, incerteza da medição, incerteza diagnóstica, período de janela, segurança pós-transfusão, serviço de sangue, valor delta.

Abstract

Post-transfusion safety of blood receptors is considered a major issue in the context of blood establishments. Considering the safety of blood receptors, these establishments must test a set of transmissible agents with a significant prevalence in the population. The role of screening tests' results in the validation of the blood donations is critical for the assurance of post-transfusion safety. In the European Union countries, this screening is required by the Directive 2001/83/EC and in US by the "standards for blood banks and transfusion services".

Due to the uncertainty associated to screening tests' results, a reported result can be false. A false positive result has no consequence in terms of the blood receptor safety. However, a false negative result has a major impact since there is a high probability of resulting in the infection of the blood component receptor. Consequently, screening laboratories should consider the evaluation of measurement uncertainty in the quality control schemes. However, the European Union or the US requests are generic when dealing with uncertainty.

This thesis collects a set of articles discussing methods for the evaluation of measurement uncertainty in screening immunoassays such as complementary approaches and models of diagnostic accuracy. The Guide to the Expression of Uncertainty in Measurement (GUM) is not systematically used in blood establishments' screening laboratories such as in other medical laboratories. However, other uncertainty estimations, namely based on statistical methods, have been used in blood establishments. Accordingly, the discussion includes not only models fulfilling GUM principles focused on uncertainty of ratio results close or equal to the "cutoff" (clinical decision value), but also complementary models: total analytical error, seronegative window period and the delta for a first estimation of the impact of uncertain results in the blood establishment budget. From the perspective of diagnostic uncertainty are reunited diagnostic accuracy models for determination of diagnostic sensitivity, diagnostic specificity and area under the receiver operating characteristic curve. Complementary are discussed models to agreement between tests' results.

To illustrate the evaluation of measurement uncertainty through the different methods, the results from a same anti-hepatitis C virus chemiluminescence immunoassay are used. The computations reveal expanded uncertainties around "cutoff" from 21 to 36 % which were used to estimate the "rejection zone" [0.70, +∞ [. Only 0.19 % of 9805 samples were in this region. The total analytical error was 23 %. Considering the immunoassay rejection zone, the window period model estimated a seronegative period of 97 days. The delta-value indicated that the infected individuals' sample tested in "test 1" is less likely to produce indeterminate results and has not been shown a statistically significant difference between "test 1" and "test 2" to generate indeterminate results in sample of healthy individuals. The diagnostic sensitivity and diagnostic specificity 95 % confidence intervals were [88.3, 100 %] and [98.5, 100 %], respectively, and the values of the area under the receiver operating characteristic curve obtained were between 0.99 and 1.00. The overall agreement of results between two anti-hepatitis C virus immunoassays was from 98 to 99 %, the positive results agreement was from 87.9 to 100 % and the negative results agreement from 97.8 to 99.9 %. The 95 % confidence interval applied to the probabilities and agreements estimates is equivalent to the expanded uncertainty concept of measurement uncertainty.

The presented methods have all application in the evaluation of measurement uncertainty and diagnostic uncertainty in an immunoassay, although under different conditions of use. A scheme for the selection of models for the estimation of uncertainty in screening laboratories in blood establishments is recommended, taking into account the role of each model in post-transfusion safety.

Key Words: blood establishment, delta-value, diagnostic accuracy, diagnostic uncertainty, GUM, measurement uncertainty, post-transfusion safety, screening immunoassay, total analytical error, window period.

ACKNOWLEDGEMENTS

The research work presented in this thesis has been carried out at the Department of Quality Management of the Portuguese Institute of Blood and Transplantation.

First of all, I would like to express my deep gratitude and thanks to Professor James O. Westgard, PhD, and Professor Pedro Encarnação, PhD, who accepted the challenge of guiding me through the PhD program. A special thanks to Dr. Gracinda de Sousa, MD, Member of the Directive Board, who made it possible for me to proceed with the research study and to Professor Helder Trindade, MD, PhD, President of the Directive Board, for the support from the Portuguese Institute of Blood and Transplantation.

I would also like to express my gratitude to all co-authors of my publications: Dr. Jerard Seghatchian, MD, PhD, Dr. Bertil Magnusson, PhD, and Professor Elvar Theodorsson, MD, PhD, for their support and valuable advice. Having them as co-authors of my papers is a great honor to me. Thanks also to Dr. Mário Pádua, MD, for being the first to motivate me to research projects in quality control in medical laboratory, and to Dr. Maria João Alpoim, MSc, for the invitation and support for a career in a blood establishment.

My appreciation goes also to Professor Isabel Vasconcelos, PhD, Dean of the Faculty of Biotechnology of the Catholic University of Portugal, for her support in the process of submitting this thesis.

Furthermore, I would like to acknowledge my near and dear family: wife Sandra Xavier and my daughter Madalena Pereira, who not only supported me, but also suffered from my absence.

Paulo Pereira

Lisbon, December 2015

This page intentionally left blank

Table of Contents

This page intentionally left blank

General Introduction

This page intentionally left blank

Project Rationale and Objectives

Background

Post-transfusion safety has a critical role in the decisions taken in a blood establishment. Examples are the quality control schemes in screening laboratories intended to assure a non-significant residual risk of post-transfusion infection. Uncertainty estimations focused on the evaluation of measurement uncertainty and assessment of risk probabilities are instrumental in screening laboratories management based on “risk-based thinking”.

However, the European Union Directives and the US standards for the consideration of measurement uncertainty in screening laboratories are generic. The selection of models for evaluating measurement uncertainty is a hard challenge since the available models do not address explicitly the requisites and particularities of blood components screening laboratories. Any measurement uncertainty model for these purposes should be focused on the results' intended use (blood component receptor's safety), must consider the uncertainty at the clinical decision point (cutoff) and its impact in the rejection zone, and must also consider the probability of false results, namely the probability of false negative results, which has a direct impact on the post-transfusion safety.

Research problem

This thesis addresses the problem of finding a suitable method to evaluate the uncertainty in screening immunoassays' results in blood establishments.

Research questions

The final goal of this work is to recommend a scheme for the evaluation of uncertainty in blood establishments' screening laboratories. To this end, the following research questions were considered:

a) Which models could be considered to the determination of measurement uncertainty and a “rejection zone” in blood establishments’ screening immunoassays fulfilling The Guide to the Expression of Uncertainty in Measurement principles?

b) Is the total analytical error an alternative approach to the determination of measurement uncertainty in screening immunoassays?

c) How should a “rejection zone” impact the definition of the seroconversion window period?

d) How to estimate the impact of uncertain results in blood establishments’ budget?

e) How should be considered the uncertainty on computed diagnostic accuracy estimates?

f) Which scheme should be recommended to select the method of uncertainty in blood establishments' screening immunoassays?

Purpose of the research study

The purpose of the study is to understand the role of uncertainty in screening laboratory’s quality management systems and its impact on post-transfusion safety, and to evaluate models for the estimation of uncertainty in screening immunoassays. An empirical study of the available models for the determination of uncertainty is performed. All methods were applied to a single screening immunoassay (antibodies to hepatitis C virus test) allowing for the direct comparison of results.

This research is limited to the routine use of commercial screening immunoassays in blood establishments. “In-house” immunoassays are outside its scope.

Outline of Thesis Structure

The thesis is divided in three parts and six chapters.

Part I introduces the questions related to the uncertainty of results in blood establishments' screening laboratories. It includes two chapters. **Chapter 1** contains a critical review of the approaches and trends in the quality management systems applied to screening laboratories. On **Chapter 2** the impact of uncertainty in the screening immunoassays' results in blood establishments is examined.

Part II, with chapters 3 to 5, contains the evaluation of the available methods to estimate uncertainty in screening immunoassays. **Chapter 3** addresses the application to screening immunoassays of measurement uncertainty modeling and empirical estimation methods fulfilling GUM principles, recommending a novel application of empirical models. The uncertainty evaluation is focused on the "cutoff", and a new utilization of the derived measurement uncertainty to the estimation of the "rejection zone" is proposed, where blood donations with indeterminate or positive test results are rejected. On **Chapter 4** the total analytical error model is reviewed as an alternative to the GUM intralaboratory empirical model. A new definition of the seroconversion window period is proposed, considering the rejection zone. Additionally, an application of the delta value for a primary evaluation of the impact of uncertain results in blood establishments' budget is presented. Uncertainty estimation through diagnostic accuracy, area under the receiver operating characteristic curve and agreement of test's results models are analyzed in **Chapter 5**. It is considered that the critical output of these models is a confidence interval for the measurement results, which is comparable to the concept of expanded uncertainty in modeling and empirical methods.

Part III includes the conclusion chapter and the suggestions for further work. **Chapter 6** contains the main contribution of this thesis: a scheme to guide the uncertainty evaluation in blood establishments' screening laboratories. This scheme is intended to support the laboratory staff to choose a set of models for evaluating measurement uncertainty and computing the rejection zone, as well as for determining the probabilities of false results and the seronegative period. General conclusion and directions for future research are presented in **Chapter 7**.

Each chapter includes the text, tables and figures of the published peer reviewed articles (chapters 1 to 6). Minor revisions have been made to the articles in order to standardize the language, notation, results presentation, tables, and figures throughout the thesis.

Paulo Pereira is the main author of these articles. Renowned scientists in the areas of medical laboratory's quality control, blood establishments' quality control and metrology scope reviewed preliminary versions of the articles and provided further guidance in the preparation of the final manuscripts. That was acknowledged by giving them the co-authorship of the articles.

Paulo Pereira's motivation to conduct this research work came from his practice as Coordinator of the Department of Quality Management of the Portuguese Institute of Blood and Transplantation. His major concerns with the impact of false negatives in the post-transfusion safety led him to demonstrate a research interest in quality control dealing with screening immunoassays' results. After a literature review he realized that there was not an established method to the determination of measurement uncertainty in screening immunoassays' results. He then worked towards the development of such measurement uncertainty evaluation method, always having in mind its application in the routine of blood establishment screening laboratories. The journal articles and conference oral and poster presentations are the major output of his research work started in September 2006. His *curriculum vitae* shows other activities and projects dealing with quality control in blood establishments, complementing his research work.

Scientific Outputs of the Thesis

Evaluating the uncertainty in blood establishments screening immunoassay's results is of major importance to assure blood components receptor's safety. However, there were no established methods to evaluate the tests results uncertainty that took under consideration the particularities of blood establishments screening immunoassays. This thesis fills that gap by critically evaluating currently available methods under the perspective of blood establishments screening laboratories, and by proposing the modifications necessary to respond the screening laboratories' needs, while taking into account the results intended use. A new scheme for the selection of the most suitable uncertainty evaluation method is proposed.

The work presented here originated several peer-reviewed journal articles and conference oral and poster presentations, which are listed below.

Publications

Peer-refereed articles in international scientific journals

Paulo Pereira, Bertil Magnusson, Elvar Theodorsson, James Westgard, Pedro Encarnação (2015). Measurement uncertainty as a tool for evaluating the "grey-zone" to reduce the false negatives in immunochemical screening of blood donors for infectious diseases. Accreditation and Quality Assurance, First Online: November 24, 2015, DOI: 10.1007/s00769-015-1180-x.

Paulo Pereira, James Westgard, Pedro Encarnação, Jerard Seghatchian, Gracinda de Sousa (2015). Quality management in the European screening laboratories in blood establishments: a view on current approaches and trends. Transfusion and Apheresis Science, 52(2):245-251, DOI: 10.1016/j.transci.2015.02.014

Paulo Pereira, James Westgard, Pedro Encarnação, Jerard Seghatchian, Gracinda de Sousa (2015). The role of uncertainty in results of screening immunoassays in blood establishments. Transfusion and Apheresis Science, 52(2):252-255, DOI: 10.1016/j.transci.2015.02.015.

Paulo Pereira, James Westgard, Pedro Encarnação, Jerard Seghatchian (2015). Evaluation of the measurement uncertainty in screening immunoassays in blood establishments: Computation of diagnostic accuracy models. Transfusion and Apheresis Science, 52(1):35-41, DOI: 10.1016/j.transci.2014.12.017.

Paulo Pereira, James Westgard, Pedro Encarnação, Jerard Seghatchian, Gracinda de Sousa (2015). Scheme for the selection of measurement uncertainty models in blood establishments' screening immunoassays. Transfusion and Apheresis Science, 52(1):42-47, DOI: 10.1016/j.transci.2014.12.018.

Paulo Pereira, James Westgard, Pedro Encarnação, Jerard Seghatchian (2014). Analytical model for calculating indeterminate results interval of screening tests, the effect on seroconversion window period:

a brief evaluation of the impact of uncertain results on the blood establishment budget. Transfusion and Apheresis Science, 51(2):126-31, DOI: 10.1016/j.transci.2014.10.004.

Peer-refereed papers in international scientific conference proceedings

Paulo Pereira, James Westgard, Pedro Encarnação, Gracinda de Sousa (2015). The receiver operating characteristic curve and area under the curve as substitute of GUM model for determination of measurement uncertainty in blood establishments' screening tests results. Oral presentation at the II International Conference on Health Research, Feb 19-21, Oliveira de Azeméis, Portugal. In Evidências, (Suplemento Fev):48-54.

Paulo Pereira, James Westgard, Pedro Encarnação, Gracinda de Sousa (2015). The role of uncertainty of transmissible diseases in post-transfusion safety. Oral presentation at the II International Conference on Health Research, Feb 19-21, Oliveira de Azeméis, Portugal. In Evidências, (Suplemento Fev):48-54.

Paulo Pereira, James Westgard, Pedro Encarnação, Gracinda de Sousa (2015). A revision of the seroconversion window period model considering the effect of measurement uncertainty on the "cutoff" point in blood establishments' screening tests. Poster presentation at the II International Conference on Health Research, Feb 19-21, Oliveira de Azeméis, Portugal. In Evidências, (Suplemento Fev):93-124.

Paulo Pereira, James Westgard, Pedro Encarnação (2014). Recommendation of measurement uncertainty models complying GUM and alternative for blood bank screening tests results. Poster presentation at the 2nd World Congress of Health Research, Oct 7-8, Viseu, Portugal. In Atención Primaria, 46(Espec Cong 1):65-87, Retrieved from http://www.elsevier.es/en-revista-atencion-primaria-27-pdf-90366858-S300.

Paulo Pereira, James Westgard, Pedro Encarnação (2014). Diagnostic models for computing measurement uncertainty in blood bank screening tests. Oral presentation at the 2nd World Congress of Health Research, Oct 7-8, Viseu, Portugal. In Atención Primaria, 46(Espec Cong 1):25-63, Retrieved from http://www.elsevier.es/en-revista-atencion-primaria-27-pdf-90366858-S300.

Paulo Pereira, James Westgard, Pedro Encarnação (2014). Practicability and relevance of GUM approach for computing measurement uncertainty in blood bank screening tests. Oral presentation at the 2nd

World Congress of Health Research, Oct 7-8, Viseu, Portugal. In Atención Primaria, 46(Espec Cong 1):25-63, Retrieved from http://www.elsevier.es/en-revista-atencion-primaria-27-pdf-90366858-S300.

Paulo Pereira, Patrícia Rodrigues, Alexandra Ribeiro, Sónia Silva, Nazaré Boavida (2007). A quality control approach applied to alanine aminotransferase assay with evidence of traceability and sigma metrics. Poster presentation at the XVIIth Regional Congress of Europe-International Society of Blood Transfusion, Jun 23-27, Madrid, Spain. In Vox Sanguinis, 93(Issue Supplement s1):54-274, DOI: 10.1111/j.1423-0410.2007.00932.x.

Communications

Oral communications in international scientific conferences

Paulo Pereira (2015). Quality management in the European screening laboratories in blood establishments: a view on current approaches and trends. Presented at the NRL European Workshop on Quality, 22nd International Workshop on Surveillance and Screening of Blood Borne Pathogens, The International Plasma Fractionation Association and the Paul-Ehrlich-Institut, May 19-21, Prague, Czech Republic.

Paulo Pereira, James Westgard, Pedro Encarnação, Gracinda de Sousa (2015). The receiver operating characteristic curve and area under the curve as substitute of GUM model for determination of measurement uncertainty in blood establishments' screening tests results. Presented at the II International Conference on Health Research, Feb 19-21, Oliveira de Azeméis, Portugal, DOI: 10.13140/2.1.3000.1445.

Paulo Pereira, James Westgard, Pedro Encarnação, Gracinda de Sousa (2015). The role of uncertainty of transmissible diseases in post-transfusion safety. Presented at the II International Conference on Health Research, Feb 19-21, Oliveira de Azeméis, Portugal, DOI: 10.13140/2.1.4048.7202.

Paulo Pereira, James Westgard, Pedro Encarnação, Gracinda de Sousa (2014). Diagnostic models for computing measurement uncertainty in blood bank screening tests. Presented at the 2nd World Congress of Health Research, Oct 7-8, Viseu, Portugal, DOI: 10.13140/2.1.2142.5922.

Paulo Pereira, James Westgard, Pedro Encarnação, Gracinda de Sousa (2014). Practicability and relevance of GUM approach for computing measurement uncertainty in blood bank screening tests. Presented at the 2nd World Congress of Health Research, Oct 7-8, Viseu, Portugal, DOI: 10.13140/2.1.4387.2003.

Oral communications in national scientific conferences

Paulo Pereira, James Westgard, Pedro Encarnação, Gracinda de Sousa (2015). The risk of incorrect clinical decisions due to the abnormal medical laboratory results. Presented at II Simpósio Ensaios Clínicos: Que Impacto na Comunidade?; Mar 20, Lisbon, Portugal.

Paulo Pereira, Pedro Encarnação (2007). Requisitos aplicados a desenhos de modelo de controlo de qualidade em laboratórios clínicos (Requirements applied to quality control designs in medical laboratories). Presented at the 2as Jornadas Internacionais da Qualidade em Laboratórios de Análises Clínicas; Nov 7-9; Lisbon, Portugal, DOI: 10.13140/2.1.2080.3365.

Paulo Pereira, Pedro Encarnação (2007). Modelo de controlo de qualidade aplicado à alanina aminotransferase com evidência de rastreabilidade e definição de métrica sigma (Quality control model applied to the alanine aminotransferase test with demonstration of metrological traceability and sigma metrics evaluation). Presented at the 2.^{as} Jornadas Internacionais da Qualidade em Laboratórios de Análises Clínicas, Nov 7-9, Lisbon, Portugal, DOI: 10.13140/2.1.2997.8408.

Paulo Pereira (2007). Gestão da qualidade em laboratório - Outros aspetos da gestão do risco em laboratórios (Quality management in medical laboratories - Alternative concepts of laboratory risk management). Presented at the 7.º Encontro de Saúde Ocupacional em Hospitais e Outros Estabelecimentos de Saúde of Sociedade Portuguesa de Medicina do Trabalho, Mar 10, Lisbon, Portugal, DOI: 10.13140/2.1.3391.0560.

Poster communications in international scientific conferences

Paulo Pereira, James Westgard, Pedro Encarnação, Gracinda de Sousa (2015). A revision of the seroconversion window period model considering the effect of measurement uncertainty on the “cutoff”

point in blood establishments' screening tests. Presented at the II International Conference on Health Research, Feb 19-21, Oliveira de Azeméis, Portugal, DOI: 10.13140/2.1.1886.0329.

Paulo Pereira, James Westgard, Pedro Encarnação (2014). Recommendation of measurement uncertainty models complying GUM and alternative for blood bank screening tests results. Presented at the 2nd World Congress of Health Research, Oct 7-8, Viseu, Portugal, DOI: 10.13140/2.1.2027.9043.

Paulo Pereira, Patrícia Rodrigues, Alexandra Ribeiro, Sónia Silva, Nazaré Boavida (2007). A quality control approach applied to alanine aminotransferase assay with evidence of traceability and sigma metrics. Presented at the XVIIth Regional Congress of Europe-International Society of Blood Transfusion, Jun 23-27, Madrid, Spain, DOI: 10.1111/j.1423-0410.2007.00932.x.

This page intentionally left blank

Part I - Introduction

This page intentionally left blank

Chapter 1 - Quality Management in the European Screening Laboratories in Blood Establishments: A View on Current Approaches and Trends

This page intentionally left blank

1.1. Abstract

The screening laboratory has a critical role in the post-transfusion safety. The success of its targets and efficiency depends on the management system used. Even though the European Union Directive 2002/98/EC requires a quality management system in blood establishments, its requirements for screening laboratories are generic. Complementary approaches are needed to implement a quality management system focused on screening laboratories.

This Chapter briefly discusses the current quality management system practices and good laboratory practices, as well as the trends in quality management system standards.

ISO 9001 is widely accepted in some European Union blood establishments as the quality management standard; however, this is not synonymous of its successful application. The ISO "risk-based thinking" is interrelated with the quality risk-management process of the EuBIS "Standards and criteria for the inspection of blood establishments". ISO 15189 should be the next step on the quality assurance of a screening laboratory, since it is focused on medical laboratory.

To standardize the quality management systems in blood establishments' screening laboratories, there is a need of new national and European claims focused on technical requirements following ISO 15189.

1.2. Introduction

Quality management is currently defined as the "management with regard to quality" (entry 3.3.4 of [1]). For its operation the medical laboratory should have a quality management system to establish policy and goals allowing targets to be achieved.

The development of a quality management process in laboratory testing started during the end of 60's by medical laboratory researchers from Norwalk Hospital, Norwalk, CT [2,3]. The total quality management (TQM) was the output of Deming and industrial approaches which has been applied to medical laboratory by researchers coming from its field [4]. It identifies components as quality laboratory processes, quality control, quality assessment, quality planning, and quality goals. These principles were applied to the management of the analytical quality of laboratory testing process (analytical quality management) [5],

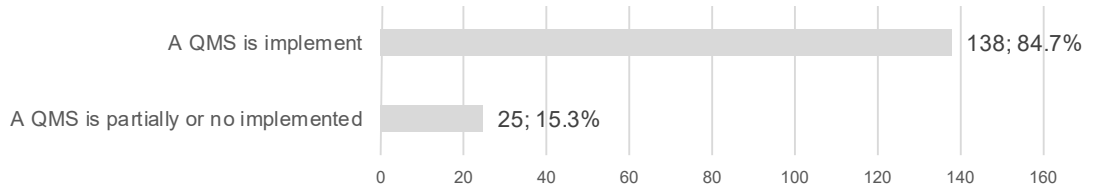
considering Deming's premise that improvements in quality lead to improvements in productivity and reductions in cost [6].

Currently, the state of quality management in medical laboratories is not comparable to that in world class industries and service organizations. It is often assumed that the high level of automated technology assures quality, but many reagent manufacturers and medical laboratories focus on minimum requirements for compliance with regulatory and inspection requirements. Quality remains an important issue in laboratory testing, even though there have been advancements in technology, improved laboratory practices, and improved quality management approaches and tools [7].

The European Union Directive 2005/62/EC requires the implementation of a management system in a blood establishment [8]. Figure 1 shows the percentage and number of European blood establishments with an implemented quality management system (QMS) from 186 blood establishments in 33 countries; the majority of the European countries had a QMS implemented in 2012. Figure 2 shows the percentage and number of quality management approaches implemented in the same sample, from which it is clear that there is a large number of different approaches whereby standardization is absent, and Figure 3 shows the percentage and number of countries where a certain quality management approach is mandatory.

Like the European Union blood establishments must have a QMS, it is expected that a major number of the European countries have the European Directives as mandatory in their national laws. The requirements applicable to the screening laboratory in Directive 2005/62/EC are generic, so the laboratory staff needs to use complementary guidance. The US Standards for Blood Banks and

Figure 1 2012 European Directorate for the Quality of Medicines & HealthCare (EDQM) Survey - Quality management system implementation in European blood establishments.



Transfusion Services of the American Association of Blood Banks [9] are mainly focused in areas of the collection and production of blood components does not requiring specific screening laboratory technical

requirements, give that they are already inspected in US according to the Clinical Laboratory Improvement Amendments (CLIA).

The major goal of a screening laboratory in a blood establishment is the post-transfusion safety assurance. It is important to review the current good management and good laboratory methodologies on medical laboratories and blood establishments' fields, since they are critical to guarantee the post-transfusion safety as well as to practice an efficient management.

Figure 2 2012 EDQM Survey - Implemented approaches of quality management system in European blood establishments.

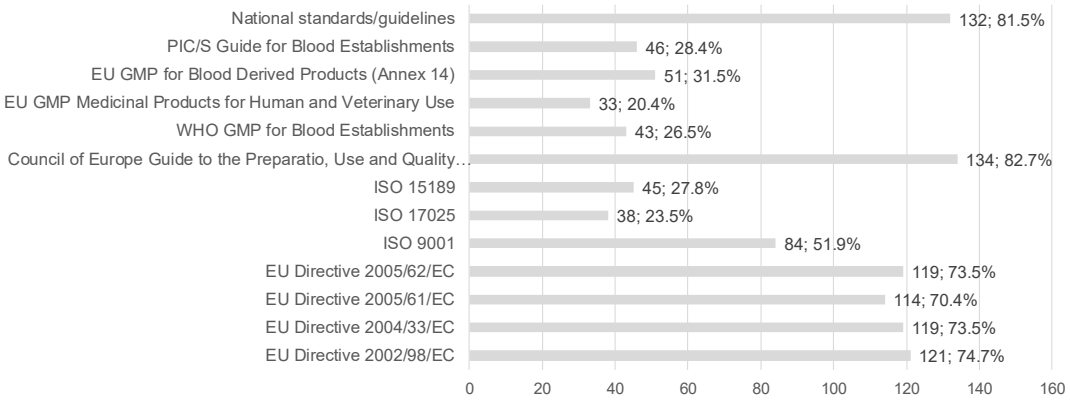
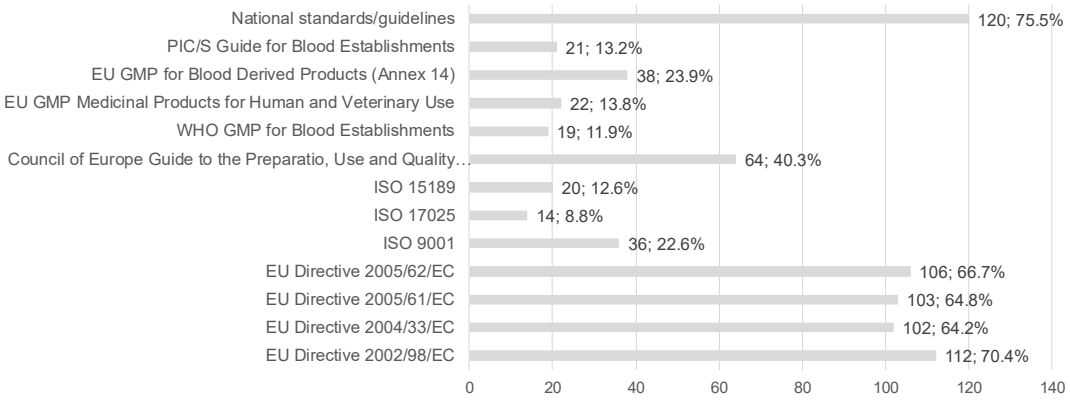


Figure 3 2012 EDQM Survey - Mandatory approaches of quality management system in European blood establishments.



1.3. Material and methods

1.3.1. Quality management system

QMS generally refers to the management and control of manufacturing according to costumer's requirements. It can be also applied to the medical laboratory processes (whole process) outputting

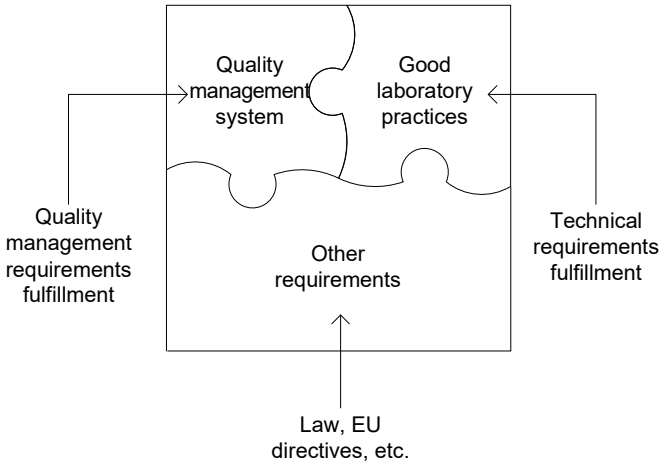
reported results. QMS promotes the adoption of a laboratory process approach when developing, implementing and improving the effectiveness of a quality management system, to enhance customer satisfaction. An important role of QMS is the documentation and records of process activities and operations involved with testing. If the documentation showing how the testing was made and tested is not correct and in order, then the reported result has a high chance to be untrue (nonconforming). Additionally, QMS requires that all testing equipment has been qualified as suitable for use, and that all operational methodologies and procedures utilized in the testing process have been validated to demonstrate that they can perform their purported function, i.e., the reported results do not compromise the clinical decision, for example, rejection of blood donation of a donor infected by hepatitis C virus post verification of screening tests' results [10-13].

The International Organization for Standardization (ISO) 9001"Quality management systems - requirements" is the global standard requesting a QMS. Its adoption should be a strategic decision of an organization [14].

1.3.2. Good laboratory practices (GLP)

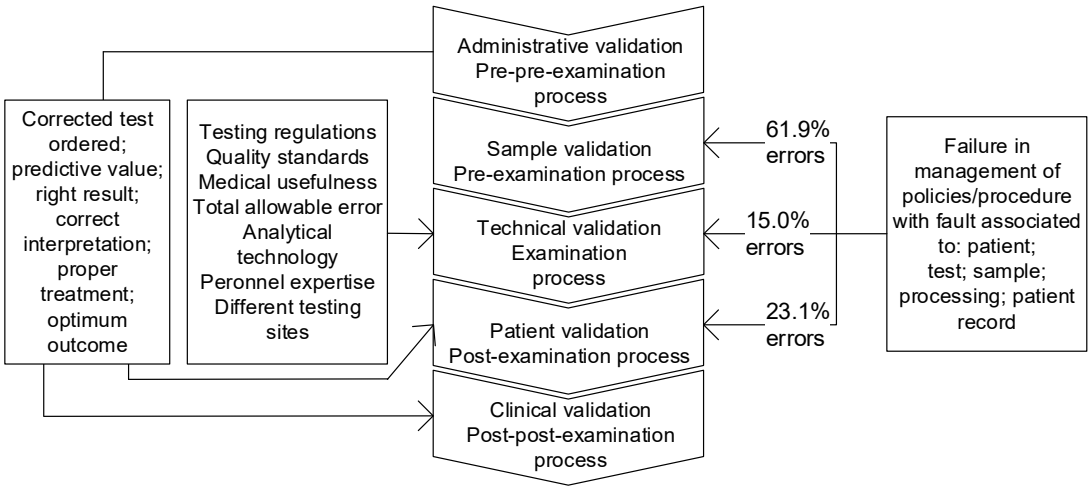
GLP generally refers to a system of management for laboratories to ensure the consistency and reliability of results as outlined in specific regulations. GLP embodies a set of principles that provides a framework within which laboratory examinations are planned, performed, monitored, recorded, reported, and archived. GLP helps assure regulatory authorities that the data submitted are a true reflection of the results obtained during the examinations and can therefore be relied upon when making risk/safety assessment. GLP is mainly associated with technical competence evidence [10,15-18]. Figure 1 shows a screening laboratory testing process scheme requiring QMS and GLP, as well as other requirements such as regulatory law. The laboratory's process in usually divided into three processes: pre-examination process or pre-analytical phase, examination process or analytical phase, and post-examination process or post-analytical phase [5,19-22]. A later model, referred as "filter model" or "NEXUS vision", proposes five processes [23]. Figure 2 shows this model with error distribution from examination process to its first related processes (pre a post-examination), and to its secondary related processes (pre-pre and post-post examination) [24]. GLP is applied in three processes where it is associated with specific laboratory technical requirements. The NEXUS pre-pre and post-post-examination processes in a hospital laboratory

Figure 1 QMS and GLP models as part of laboratory requirements set.



deal mainly with physicians' events related to test requests and associated actions, such as collection, and to the use of test results for patient treatment, which must be according to patients' interest. In blood establishment the pre-pre and post-post examination process are legislated in European Union countries according to regulatory Directives and national law.

Figure 2 Testing processing flow, interactions according to NEXUS vision, and error rates from Carraro and Plebani (2007) with error causes by Ehrmeyer and Laessing (2006).



ISO 15189 is the standard for medical laboratory tests, focused principally in GLP [22]. It is mandatary in Australia, Latvia and in France after November 1, 2016. Outside of ISO's global guidelines, there are a few national GLP programs and requirements except for CLIA (US), the National Association of Testing Authorities (NATA) (Australia), CCKL (The Netherlands), and RiliBÄK (Germany).

1.3.3. ISO 9001

ISO 9001 is intended for certification of organizations of all sizes and types, which explains why it is a generic approach. According to ISO standard, the fulfillment of testing requirements should guarantee interested parties' satisfaction including regulatory compliance. It is important to consider all interested parties (for example, a national agency for the regulation of blood establishments) and not uniquely the costumer. ISO 9001 requirements are focused on the quality improvement cycle referred as PDCA (plan-do-check-act), providing a logical and scientific management model for continuous quality improvement. Current edition requires "risk based thinking" in decision-making happening in the four stages of PDCA cycle. "Risk" is in defined as "the effect of uncertainty" (entry 3.7.9 of [1]), i.e., risk is equivalent to a deviation from the expected result. ISO 9001 is applicable to the whole organization, but does not consider specific/technical laboratory requirements. Despite being very well known in manufacturing industries; ISO standards are relatively new in blood establishments. As the certification of blood establishments has become more widespread, its applicability to blood establishments has been better understood.

The PDCA approach enables the screening laboratory to improve the quality of its reported results, to guarantee the continuous satisfaction of its customers who may be clinical decision in patients, post-transfusion safety, post-transplantation safety or others.

The practical approach to meet ISO 9001 requirements must be designed at each medical laboratory, and there is not any "single approach" to be used. For ISO 9001 to be successfully applied, there must be a dynamic PDCA guaranteeing continual improvement and customer satisfaction. The PDCA starts with the responsibility of top management to provide evidence of its commitment to the development and implementation of the quality management system and to continually improve its effectiveness. Top management must define which plans/projects are needed to achieve defined targets, which should be based on intended use and customer satisfaction. The need of human resources, infrastructure, and work environment must be identified and should be guaranteed by top management. Product of realization is the laboratory reported results, which should guarantee customer's satisfaction. Production processes must be controlled and periodically monitored by appropriate quality indicators, as well as periodic internal audits, control of nonconforming results, implementation of corrective and preventive actions,

and verification of the achievement of the intended quality of results and services. Laboratories must define their methodology to measure customer satisfaction. The evaluation of customer satisfaction should be part of management review, as well as other indicators that are specified in ISO's requirements for management review. After review, top management should take actions for continual improvement. Patients have limited ability to evaluate their satisfaction (final auditor), but a "perfect" reported result is the one that does not limit the clinical decision due to any laboratory process nonconformity [14].

The success of ISO 9001 dynamics in screening laboratory depends principally on how are used the seven principles of quality [1]:

a) Customer-focus - Customers' requirements must be defined; they may be defined by state laws, but they are not synonymous of technical requirements; laboratories can define non-regulatory requirements by contracts of agreement;

b) Leadership - Unity of purpose and direction should be established to create an internal environment where all the staff can contribute to the achievement of the laboratory's goals;

c) Engagement of people - All the staff should be fully involved to enable their abilities to be used for the benefits of blood establishment;

d) Process approach - The processes must be identified and managed, defining their interactions in order to achieve an efficient flow and produce acceptable results and the interrelation of all processes must be managed to guarantee efficiency;

e) Improvement - It should be a permanent objective to continuously improvement performance and quality;

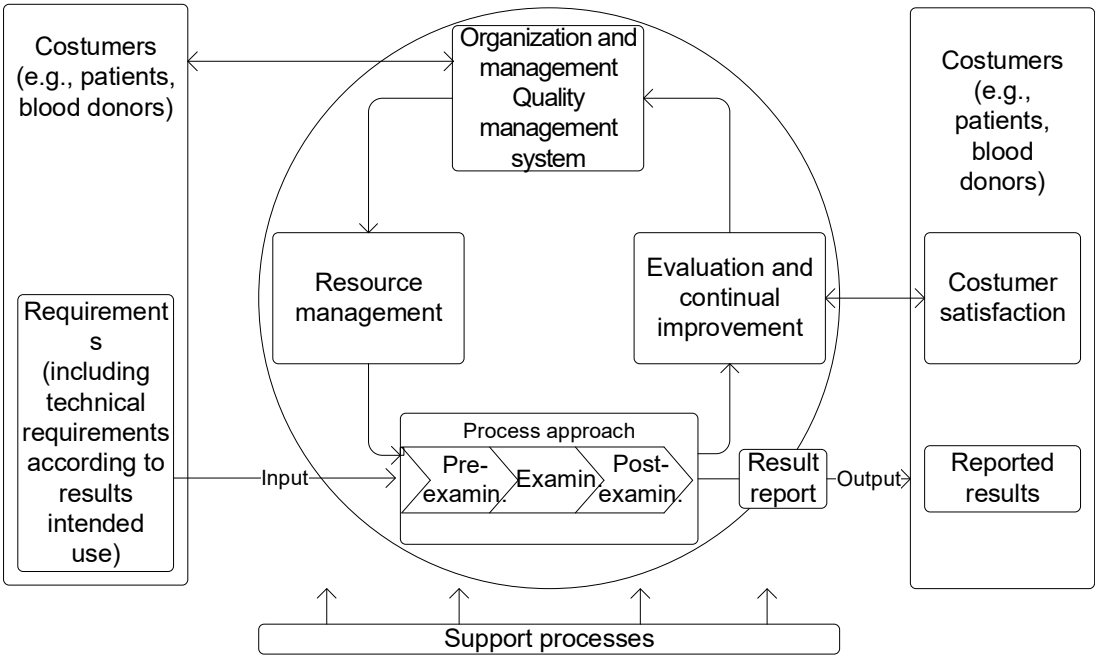
f) Evidence-based decision making - Decisions must be based on logical analysis and reliable data;

g) Relationship management - The relationship with interested parties, such as suppliers (agreement), should enhance the laboratory and suppliers' ability to create value; the laboratory must define and inform suppliers about their purchase requirements for products or services.

1.3.4. ISO 15189

The first edition of ISO 15189 was published in 2003 and was the first global guideline to focus specifically on medical laboratory technical requirements, adapting many of the requirements of ISO/IEC 17025 for methods of testing and calibration. Its purpose is to guarantee the satisfaction of medical laboratory costumers [22]. Its scope is uniquely the medical laboratory (including those at hospitals, blood, cells and tissues establishments, sports agencies, etc.).

Figure 3 Continual improvement of the quality management system according to ISO 15189 in a blood establishment laboratory.



Its process approach follows the three process model where (a) the pre-examination processes "in chronological order, from the clinician's request and include the examination request, preparation and identification of the patient, collection of the primary sample(s), and transportation to and within the laboratory, and end when the analytical examination begins", (b) the examination processes are the "set of operations having the object of determining the value or characteristics of a property", and (c) the post-examination processes comprises the "processes following the examination including review of results, retention and storage of clinical material, sample (and waste) disposal, and formatting, releasing, reporting and retention of examination results". Figure 3 illustrates the TQM cycle applied to a blood

establishment laboratory with tests accredited by ISO 15189 covering the three processes. Seghatchian and Stivala (1992) had already encouraged the use of the TQM cycle in blood establishments [25].

ISO 15189 features a set of detailed requests including requirements for personnel, accommodation and environmental conditions, laboratory equipment, reagents, and consumables, pre-examination processes, examination processes, ensuring quality of examination processes results, post-examination processes, reporting of results, release of results, and laboratory information management.

For equipment, the laboratory must document the procedure for calibration of equipment affecting directly or indirectly the reported results. The guideline states the minimum requirements, summarizing: it should consider all of manufacturer's instructions commonly in the reagent insert paper and in operation manual, data allowing metrological traceability and calibration and recalibration status/post-actions, periodic verification of measurement accuracy, and adjustments or other interfering actions that changes the equipment condition.

For metrological traceability, screening tests do not have available reference materials or reference procedure of the higher metrological order. The provision of confidence in the results is according to the samples of patients and healthy individuals as clearly recognized, identified, described and agreed in medical laboratory [26].

For quality control, the guideline requires the tests should be:

a) **Selected** to satisfy its intended use. ISO 15189 does not recommend any approach for selection of a new test, it does not state any approach for method validation, and it does not state any specifications for method validation;

b) **Verified** to document the performance stated by the manufacturer. The verification must evidence that the laboratory performance claims have been met. The performance specifications for each test must consider the intended use of outputted results. The verification must evidence the manufacturer claimed performance happens in the screening laboratory;

c) **Validated** to demonstrate the performance required for the intended clinical purpose. All validation specifications must be according to test or method intended use. There are uniquely required to be

validated the “non-standard methods”, “laboratory designed or developed methods”, “standard methods used outside their intended scope”, and “validated methods subsequently modified”. The validation should include diagnostic specificity and diagnostic sensitivity. The practical question is whether the screening tests should be verified or validated? To verify the manufacturer performance is possible, but large samples are not practical in small blood establishments. To validate the test could also be required for example when the blood donors and the manufacturer samples do not match, or when matching the 95 % confidence interval is required and it is unavailable in the manufacturer’s literature. Verification or validation could share the same protocol for measurement of diagnostic sensitivity and diagnostic specificity. When a test is modified it must be reevaluated.

d) **Measurement uncertainty** must be determined, even though there is no recommended approach. VIM defines measurement uncertainty as “non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used” (entry 2.26 of [27]). ISO recommends GUM principles for its evaluation. However, there are also alternative approaches that are acceptable. Given there are always uncertain results in screening tests, results close or equal to the “cutoff” have a high chance to be classified as false. So, the screening laboratory should evaluate the expanded uncertainty (entry 2.35 of [27]) and evaluate the uncertainty interval (“grey-zone”). Results in “grey-zone” area should be classified as indeterminate and further testing should be required to assure correct classification (see Chapter 3). Alternatively, the uncertainty could be equivalent to the chance of false results in a 2x2 contingency table (diagnostic uncertainty) (see Chapter 5). The performance requirements for the measurement uncertainty, such as for selection, verification and validation, must be defined and the determination should happen periodically. The laboratory must consider measurement uncertainty when interpreting measured quantity values [27].

To assure the quality of the reported results, the laboratory must have a written internal quality control (IQC) procedure to verify the accuracy of the results. Results out-of-control must be rejected and the cause, correction, corrective and preventive actions deployed.

The IQC material should have a concentration close to the clinical decision value, i.e., “cutoff”, regardless the binary result to be positive or negative. Currently, most of the materials concentrations in screening laboratories do not fulfill this principle. The laboratory must also participate in proficiency testing schemes

to evaluate laboratories' results against the measurements taken from the group of comparison program. Such as in IQC, causes and actions must be taken for nonconforming results [23].

ISO requires that all approaches in the examination process must consider the results' intended use. Intended use of a result is the ultimate role of a result in clinical decision, blood for transfusion, cell or tissue to transplantation, etc. The intended use of a result in screening laboratories in blood establishments is the post-transfusion safety. The selection of any examination process must consider this intended use. For example, the method validation of a screening test in a blood establishment or in a hospital medical laboratory does not have the same request, for example, a false negative result has a higher probability of damage to a transfused blood unit in a patient, and otherwise in the hospital laboratory there is no risk of transmission related to transfusion.

1.4. Results and discussion

ISO 9001 does not include any technical requirements for quality and competency. A limitation of ISO certification is that it does not standardize technical requirements, allowing different levels of quality per laboratory, even when the same testing services are offered. Another limitation is that does not have efficiency requirements; in a worst case, a laboratory could report results with a low or high chance to be false and be financially unsustainable. The sustainability is a limitation common to ISO 15189. By itself, ISO 9001 does not guarantee that reported result is the output of a technical evidence-based test.

In contrast, ISO 15189 requires a detailed set of technical requirements. Its implementation complements most of the regulatory requirements in different countries. The laboratory must select a set of protocols for the examination process. These should be published in recognized textbooks, peer-refereed journal articles, international or national consensus texts or laws. The Clinical and Laboratory Standards Institute (CLSI) is a global organization that provides a large and diverse number of consensual protocols for medical laboratories [28]. Its guidelines are the reference literature for medical laboratory as well as to the manufacturer. The authors are researchers with recognized experience in different fields coming from laboratories, manufacturers and regulatory agencies, and final drafts have a public review. Two of the CLSI references documents for the screening tests are the EP12-A2 intended for the evaluation of binary

results tests and EP24-A2 intended for the assessment of the diagnostic accuracy of using the receiver operating characteristic curves.

One way to assess the status of quality management in medical laboratories is to review the literature on nonconformities. The effect of failures in medical laboratories has been reported. Published articles place the nonconformities between 62.4-86.7 % in both pre and post examination processes, and 13.3-31.6 % in analytical level [29-34]. A recent study shows that the relative frequency of nonconformities is still close to that observed in 1996 (85 % for both pre and post analytical, and 15 % for analytical), but with a lesser absolute frequency amount [35]. These studies are sometimes misunderstood and interpreted to mean that the quality of the examination process is no longer important. All nonconformities should be interpreted and analytic errors are often difficult to recognize. Some studies depend on recognition of nonconformities by physicians and nurses (post-post examination), which bring serious limitation to the study. In most studies there is no information about laboratory quality management policies and the quality requirements for tests (for example, allowable total analytical error), which is critical to understand nonconformities management. For a laboratory with a quality management approach according to ISO 9001 only, there may be little or no attention paid to determining technical quality requirements and nonconformities rates. For a laboratory with tests accredited by ISO 15189, there should be a better sensitivity to detect technical nonconformities, as well a better capacity to have successful corrective and preventive actions. A technical requirement failure in ISO 15189 might not be a failure in ISO 9001 scope. A laboratory which does not follow best policies applied to analytical area might not detect some non-conformity, and has limited capability to search technical failures causes and to develop successful corrective actions. It is also important to understand that the amount of failures is not a clear principle for nonconformities management, and the core principles should be sensitivity and efficiency addressed. Nonconformities data must also be considered according to their impact (risk) on the customer (patient).

1.5. Conclusion

ISO 9001 is a standard for a quality management system and the current edition introduces a risk based approach adapting to current practices in industry and services. However, the screening laboratory should understand limitations of ISO 9001, particularly the lack of technical requirements. ISO 15189 represents

a next step in a quality management system orientated to assure the conformity of the report results. The screening laboratory staff should select protocols for the examination process considering the results intended use. CLSI protocols are recommended since they are recognized in the medical laboratories/screening laboratories and satisfy ISO 15189 requirements.

This page intentionally left blank

References

1. International Organization for Standardization. ISO 9000 Quality management systems - Fundamentals and vocabulary. 4th ed. Geneva: ISO, 2015.

2. Barnett R, Youden W. A revised scheme for the comparison of quantitative methods. Am J Clin Pathol 1970, 54(3):454-462.

3. Barnett R. Medical significance of laboratory results. Am J Clin Pathol (1968), 50(6):674-676.

4. Westgard J, Burnett R, Bowers G. Quality management science in clinical chemistry: A dynamic framework for continuous improvement of quality. Clin Chem 1990, 36(10):1712-1716.

5. Westgard J, Barry P. Cost-effective quality control: Managing the quality and productivity of analytical processes. Washington (DC): AACC Press, 1986

6. Deming W. Quality, productivity, and competence position. Cambridge (MA): Massachusetts Institute of Technology, Center for Advanced Study, 1982.

7. Westgard J. Six sigma quality design and control, 2nd ed. Madison (WI): Westgard QC, 2006.

8. Commission Directive 2005/62/EC Implementing Directive 2002/98/EC of the European Parliament and of the Council as regards Community standards and specifications relating to a quality system for blood establishments. Official Journal of the European Union L256/41-48.

9. American Association of Blood Banks. Standards for blood banks and transfusion services. 29th ed. Bethesda (MD): AABB, 2014.

10. Carston P. Good clinical, laboratory and manufacturing practices: Techniques for the QA professional. Cambridge: Royal Society of Chemistry, 2007.

11. Bunn G. Good manufacturing practices for pharmaceuticals. 6th ed. New York (NY): Informa HealthCare, 2006.

12. Sharp J. Good pharmaceutical manufacturing practice: Rationale and compliance. New York (NY): Informa HealthCare, 2004.

13. World Health Organization. Good manufacturing practices and inspection (quality assurance of pharmaceuticals). Geneva: WHO, 2007.

14. International Organization for Standardization. ISO 9001 Quality management systems - Requirements. 5th ed. Geneva: ISO, 2015.

15. Seiler J. Good laboratory practice: The why and the how. 2nd ed. Berlin: Springer, 2005.

16. Weinberg S. Good laboratory practice regulations (Drugs and pharmaceutical sciences). 4th ed. New York (NY): Informa HealthCare, 2007.

17. Anderson M. GLP essentials: A concise guide to good laboratory practice. 2nd ed. New York (NY): Informa HealthCare, 2002.

18. Garner W, Bare M, Ussary J. Good laboratory practice standards: Applications for field and laboratory studies. Washington (DC): American Chemical Society, 1992.

19. Galen R, Gambino S. Beyond normality: The predictive values and efficiency of medical diagnoses. New York (NY): Wiley & Sons Press, 1975.

20. Truchaud A, Le Neel T, Brochard M, Malvaux S, Moyon M, Cazaubiel M. New tools for labortory design and management. Clin Chem 1997, 43(9):1709-1715.

21. Burnett D. Practical guide to accreditation in laboratory medicine. Washington (DC): AACC Press, 1997.

22. International Organization for Standardization. ISO 15189 Medical laboratories - Particular requirements for quality and competence. 2nd ed. Geneva: ISO, 2007.

23. Goldschmidt H. The NEXUS vision. An alternative to the reference value concept. Clin Chem Lab Med 2004, 42(7):868-873.

24. Ehrmeyer S, Laessing R. Regulatory compliance vs. NEXUS quality for laboratory tests. Accred Qual Assur 2006, 11(6):265-268.

25. Seghatchian J, Stivala JFA. Conceptual acceptance of quality assurance and the implementation of total quality management in blood transfusion services. In: Quality Assurance in Transfusion Medicine. Rock G, Seghatchian J (editors). Boca Raton (FL): CRC Press, 1992; Vol. 1; p 2-31.

26. Clinical and Laboratory Standards Institute. EP12-A2 User protocol for evaluation of qualitative test performance. 2nd ed. Wayne (PA): CLSI, 2008.

27. Bureau International des Poids et Mesures. JCGM 200:2012 International vocabulary of metrology - Basic and general concepts and associated terms, 2008 version with minor corrections. 3rd ed. Sèvres: BIPM, 2012. Retrieved from: http://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf Accessed December 22, 2015.

28. Clinical and Laboratory Standards Institute. Retrieved from: http://clsi.org/ Accessed December 22, 2015.

29. Lapworth R, Teal T. Laboratory blunders revisited. Ann Clin Biochem 1994, 31(1):78-84.

30. Goldschmidt H, Lent R. Gross errors and work flow analysis in the chemical laboratory. Klin Biochem Metab, 1995, 3:131-140.

31. Nutting P, Main D, Fischer P, Stull T, Pontious M, Seifert M, et al. Problems in laboratory testing in primary care. JAMA 1996, 275(8):635-639.

32. Plebani M, Carraro P. Mistakes in a stat laboratory: Types and frequencies. Clin Chem 1997, 43(8):1348-1351.

33. Stahl M, Lund E, Brandslund I. Reasons to a laboratory's inability to report results for requested analytical tests. Clin Chem 1998, 44(10):2195-2197.

34. Hofgärtner W, Tait J. Frequency of problems during clinical molecular-genetic testing. Am J Clin Pathol 1999, 112(1):14-21.

35. Carrato P, Plebani M. Errors in stat laboratory: Changes in type and frequency since 1996. Clin Chem 2007, 53(7):1-5.

This page intentionally left blank

Chapter 2 - The Role of Uncertainty in Results of Screening Immunoassays in Blood Establishments

This page intentionally left blank

2.1. Abstract

The risk of uncertain results in infectious agents' tests is recognized in blood establishments, being particularly evident during the blood donor selection and in laboratory screening tests. The current risk-based approaches require risk assessment and "risk-based thinking". Accordingly, the blood establishment should consider the effect of uncertainty in all the technical decisions taken. Since the post-transfusion safety is one of the major blood establishments' goals, the risk of post-transfusion infection should be evaluated and actions taken to decrease the chance of blood donations validation use false negative results. This Chapter reviews and discusses the sources of uncertainty of infectious agents' reported results in blood establishments. It presents a scheme describing a set of sources of uncertainty that should be considered in the decisions of screening immunoassays'. The infectious agents' uncertainty concern is critical for reporting reliable results.

2.2. Introduction

The role of uncertain results is already stated in blood processing literature [1]. Uncertain results are those with a statistical significant probability to be false. Decisions taken on these results have a high chance of being incorrect. During the production process of blood components there is always a chance of uncertain results. They must be identified and actions must be taken. In blood establishments the chance of uncertain results due to the seroconversion window period was recognized principally since the first events of post-transfusion infection caused by human immunodeficiency virus (HIV) [2]. To decrease the probability of seronegative donations of infected donors, a set of questions are used to survey and select blood donors. However, there are other causes of uncertain screening tests' results.

The European Union Directives [3] and the US Standards [4] do not require the determination of uncertainty in screening tests' results. They are mainly focused in blood collection and processing. The laboratory technical requirements should include complementary approaches to evaluate uncertainty [note: in US the laboratories' test in blood establishments must fulfill the Clinical Laboratory Improvement Amendments (CLIA)] [5]. EuBIS "Standards and criteria" [6] recommend risk assessment, i.e., the evaluation of the effect of uncertainty. Whether or not required, the blood establishments must be

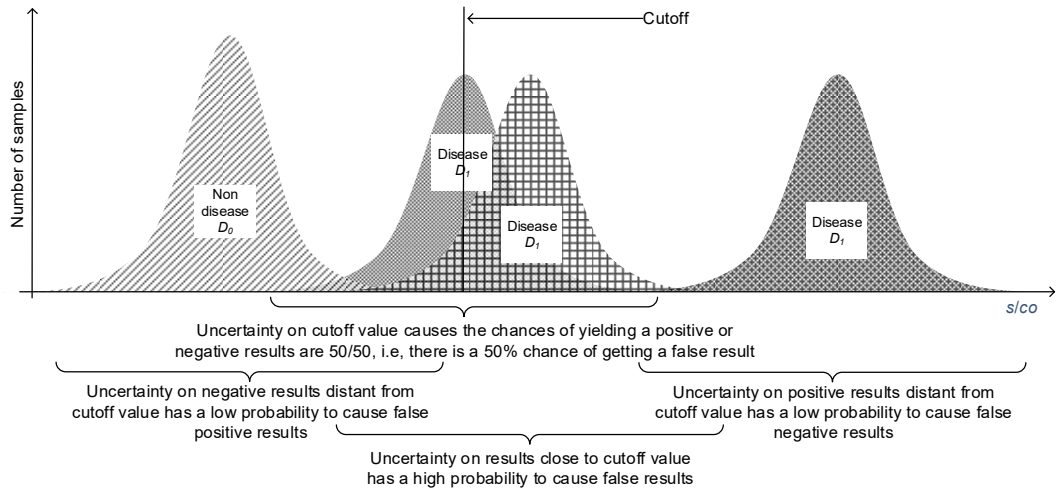
focused on the post-transfusion safety, and must evaluate the uncertainty in the screening tests' decision (for example, test selection, verification, validation, internal quality control, external quality assessment) to report reliable results. This Chapter reviews and discusses the sources of uncertainty in screening immunoassays' results, where the risk of false negatives must be considered.

2.3. Material and methods

2.3.1. Measurement uncertainty at the cutoff: "grey-zone"

The "cutoff" is the clinical decision value in screening immunoassays. The numerical results for samples are classified on an ordinal scale (entry 1.26 of [7]) according to cutoff. The results could be binary, positive or negative, or ternary, positive, indeterminate or negative. Considering the ratio between the sample result and the "cutoff" result, the "cutoff" is defined as one. Numerical values equal or close to the "cutoff" have a significant chance to be false. Therefore, the measurement uncertainty (entry 2.26 of [7]) at the "cutoff" value should be determined (see Chapter 3). The interval of uncertain results, i.e., expanded

Figure 1 Risk of false negative results in screening tests according to the value in the ordinal scale.



uncertainty (entry 2.35 of [7]) is designated as "grey-zone". Some authors determine this interval using an alternative to the measurement uncertainty: the total analytical error model [8] (see Chapter 4). Consequently, the laboratory should apply a ternary classification to the numerical results. Figure 1 shows a theoretical risk estimate of disease sampling with average equal to the "cutoff", close to the "cutoff"

and far from the cutoff; the results of infected individuals equal or close to the "cutoff" have a higher significant probability to be false negatives.

Blood donations of donors with indeterminate results must be confirmed whenever a positive result occurs. Despite the number of "grey-zone" results is relatively small to the number of blood collections (for example, the indeterminate results in the Laboratory of Transmissible Agents of the Blood and Transplantation Center of Lisbon considering a "grey-zone" equal to ± 20 %, was just 0.24 % in a sample of 9805 donors' specimens tested from September 12 to December 21, 2014), the post-transfusion safety can fail [9].

2.3.2. Uncertain diagnostic accuracy estimations: sensitivity, specificity, and area under receiver operating characteristic curve

Diagnostic sensitivity and diagnostic specificity are required to validate test performance when the output is a binary nominal quantity (entry 4.6 of [7]) result. Both are measured as a ratio, usually using a percentage, of the number of true positive results in the infected individuals' sample and the number of true negative results in the healthy individuals' sample, respectively. Other characteristics that may be useful are the ratio of the false negative results in the infected individuals' sample, and the number of false positive results in the healthy individuals' sample, *ß*-error and *α*-error, respectively. The false results are biased results (entry 2.18 of [7]). Other ratios could be determined, such as predictive values and specificity; however, their role according to results intended use in blood establishments is not significant. All these determinations are from two samples, and cannot be inferred to the populations of infected and healthy individuals. However, this inference could be done to a 95 % confidence interval (95 % CI) (or other confidence level). The reliability of the interval is linked to the number of samples and the number of true results. The inference to the infected individuals' population is to one with the same biological characteristics as the sampling used, such as types of virus [10,11]. A complementary determination to diagnostic sensitivity and diagnostic specificity is the area under the receiver operating characteristic curve. It measures the ability of the test to correctly classify positive and negative samples, often called discrimination. It ranges from 0.5, for a test that randomly assigns a positive or negative result to a sample, to 1, for a perfect test (sensitivity = 1 and 1 – specificity = 0). An estimate of uncertainty could be

determined for the area, considering a 95 % CI, representing the uncertainty associated to the capacity to discriminate true results [12]. For further information about diagnostic accuracy see Chapter 4.

2.3.3. Biological biased results: Seroconversion window period

The seroconversion window period or seronegative period is one of the primary sources of infectious agents' false results. Bhushan & Vasan (2009) defined the window period as "(…) the time between first infection and when the test can reliably detect that infection", being "dependent on the time taken for seroconversion" [13]. Fauci and Lane (2008) presented a simpler definition, "the interval between infection and detection" [14]. The determination requires a seroconversion panel(s). Since the panel is specific for an infected individual, the observed seronegative period cannot be inferred to the population of blood donors. Indeed, the true window period is unknown. For further information about window period see Chapter 5.

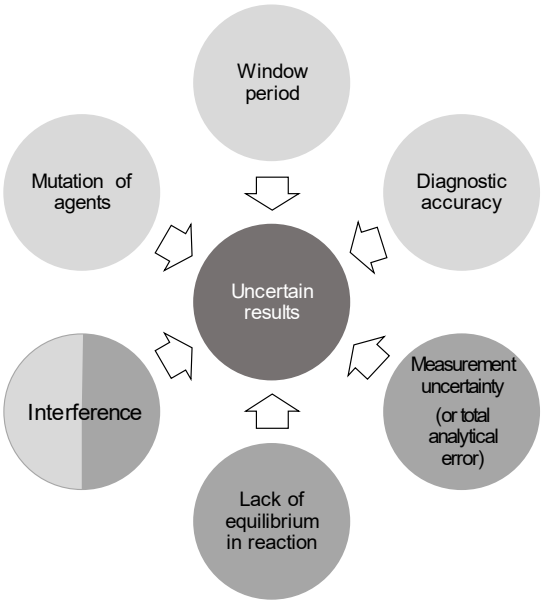
2.3.4. Biased results caused by interferences

Biased results could arise also from interfering factors. The effect of these factors is a specimen-dependent bias. Potentially interfering substances must be assessed during the selection and evaluation of a method. Young (2000-2007) published a set of databases of reported effects of pre-analytical variables, disease, drugs and herbs and natural products in the medical laboratory results [15-18]. The pre-analytical effects must be addressed by good laboratory practices. If the sampling, centrifugation and storage conditions meet the requirements stated in the manufacturer's directions for use, the chance of interferences should be low. If the laboratory practices do not fulfill these requirements, the chance of false results increases. For example, a hemolysed sample may affect the chemical or physical properties of sample matrix. The reagent kit insert should also disclose any interferences and the laboratory should verify also this information. Despite the fact that blood donors are carefully selected, there are still a set of conditions that do not cause rejection of the donation, but that could still cause to interferences, for example, rheumatoid arthritis. Also some drugs that are used, but not declared in blood donor selection, could also cause interferences, for example, drugs of abuse.

2.3.5. Lack in the equilibrium of immunoassay reaction

Poor quality laboratory practices, such as incorrect preparation, handling and storage of reagents, and inadequate reaction conditions may not guarantee the equilibrium of test reaction. The result will have a high probability to be false. The lack of equilibrium, when statistically significant, could be shown as measurement uncertainty (entry 2.26 of [7]) component [19]. The effect of nonconforming reaction conditions may be detected by the analyzers calibration verification process and should also be detected in an internal quality control scheme. Figure 2 shows a scheme that links the explained sources of uncertainty and their effect in the results.

Figure 2 Sources of uncertain results



2.4. Results and discussion

Currently, the determination of measurement uncertainty is not generally applied in blood establishments' screening immunoassays. Some of the reasons are that the blood establishments' standards and legislation does not require it, the complementary medical laboratory standard does not focus on measurement uncertainty [20], the difficulties in estimating uncertainty by modelling and empirical determinations [21,22], and the metrology vocabulary [7] is not systematically used in blood establishments.

The screening laboratory should evaluate the uncertain results by determining the window period, interference factors if recognized, and evaluating the “grey-zone”, and reporting diagnostic accuracy measurements considering a 95 % CI for reliable results.

The effect of measurement uncertainty in screening immunoassays is also estimated at the manufacturer level, where sometimes the total analytical error model is used alternatively [8]. Theoretically, the constant value in the “cutoff” equation must assure a lower chance of false negative results compared to the chance of false positive results, when a test is intended for a blood establishment. The information about “cutoff” selection by the manufacturer is not generally provided of available to the laboratory. Furthermore, screening laboratories do not verify the “cutoff” value using, for example the receiver operating characteristic curve [11,23], since it is not practical. The national agencies for the regulation of *in vitro* medical devices, have the role to accept or to reject the manufacturers’ validations, but laboratories should also make a critical assessment of performance and uncertainty. In the laboratory, a “grey-zone” should be considered and the measurement uncertainty should be determined to evaluate the zone [24,25].

Since the uncertainty affects the reliability of test results, the screening laboratory must have an internal quality control scheme to verify the conformity of control material results in analytical runs (entry 7.1 of [26]). The errors observed with control materials are linked to the errors on blood donations’ samples. Internal quality control results that are out-of-control indicate a high risk that the blood donor sample results may be incorrect. Westgard et al. (1981) published a set of rules that should be selected giving the risk to produce nonconforming results [27]. The rules are selected considering the test bias, precision (entry 2.15 of [7]) and the allowable (total analytical) error in a sigma-metric [26]. Even when a screening test is monitored by a set of rules according to its performance, there is still uncertainty associated with the control decision. The reliability of an internal quality control scheme can be described by the chance of a result with unacceptable error being classified as out-of-control (rejected), called probability of error detection, and the chance of a result with acceptable error to be classified as in-control (accepted), called probability of false rejection. Further information about internal quality control can be found elsewhere [26].

2.5. Conclusions

The uncertainty of results of screening tests is one of the major factors of post-transfusion infection. The screening laboratory should understand the causes and sources of uncertain results and how they may be estimated or determined in the laboratory. The estimation of the diagnostic accuracy models 95 % CI is critical for a consistent evaluation. A reliable internal quality control scheme should happen to guarantee a small risk of report false negative results.

This page intentionally left blank

References

1. Seghatchian J, Stivala JFA. Conceptual acceptance of quality assurance and the implementation of total quality management in blood transfusion services. In: Quality Assurance in Transfusion Medicine. Rock G, Seghatchian J, eds. Boca Raton (FL): CRC Press, 1992; Vol. 1; p 2-31.

2. Cooper DA, Gold J, Maclean P, Donovan B, Finlayson R, Barnes TG, et al. Acute AIDS retrovirus infection. Definition of a clinical illness associated with seroconversion. Lancet 1985, 1(8428):537-540.

3. Commission Directive 2005/62/EC Implementing Directive 2002/98/EC of the European Parliament and of the Council as regards Community standards and specifications relating to a quality system for blood establishments. Official Journal of the European Union L256/41-48.

4. American Association of Blood Banks. Standards for blood banks and transfusion services. 29th ed. Bethesda (MD): AABB, 2014.

5. Clinical Laboratory Improvement Amendments. CLIA law & regulations. Retrieved from: http://wwwn.cdc.gov/clia/Regulatory/default.aspx Accessed: December 22, 2015

6. European Blood Inspection System. Common European standards and criteria for the inspection of blood establishments. The European Union: EuBIS, 2010.

7. Bureau International des Poids et Mesures. JCGM 200:2012 International vocabulary of metrology - Basic and general concepts and associated terms, 2008 version with minor corrections. 3rd ed. Sèvres: BIPM, 2012.

8. Westgard J, Neil Carey R, Wold S. Criteria for judging precision and accuracy in method development and evaluation. Clin Chem 1974, 20(7):825-833.

9. Müller B, Nübling C, Kress J, Roth W, De Zolt S, Pichl L. How safe is safe: new human immunodeficiency virus type 1 variants missed by nucleic acid testing. Transfusion 2013, 53(10): 2422-2430.

10. Clinical and Laboratory Standards Institute. EP12-A2 User protocol for evaluation of qualitative test performance. 2nd ed. Wayne (PA): CLSI, 2008.

11. Pereira P, Westgard J, Encarnação P, De Sousa G. Diagnostic models for computing measurement uncertainty in blood bank screening tests. Proceedings of the 2nd World Congress of Health Research, Oct 7-8, 2014; Viseu, Portugal. In: Atención Primaria, 46(Espec Cong 1):49-50.

12. Clinical and Laboratory Standards Institute. EP24-A2 Assessment of the clinical accuracy of laboratory tests using receiver operating characteristic curves. 2nd ed. Wayne (PA): CLSI, 2011.

13. Le, T., Bhushan, V., & Vasan, N. First aid for the USMLE step 1. 20th ed. New York (NY): McGraw-Hill Medical, 2009.

14. Fauci A, Clifford Lane, H. Infectious Diseases, Human Immunodeficiency Virus Disease: AIDS and Related Disorders. In: A. Fauci, D. Kasper, D. Longo, E. Braunwald, S. Hauser, J. Jameson, J. Loscalzo (editors). Harrison's principles of internal medicine. 17th ed. New York (NY): McGraw-Hill Medical, 2008; p. 1164.

15. Young D. Effects of drugs on clinical laboratory tests. Volumes 1 & 2. 5th ed. Washington (DC): AACC Press, 2000.

16. Young D. Effects of Disease on Clinical Laboratory Tests. Volumes 1 & 2. 4th ed. Washington (DC): AACC Press, 2001.

17. Young D. Effects of Preanalytical Variables on Clinical Laboratory Tests. 3rd ed. Washington (DC): AACC Press, 2007.

18. Narayanan S, Young D. Effects of Herbs and Natural Products on Clinical Laboratory Tests. Washington (DC): AACC Press, 2007.

19. Weeks I, Beheshti I, McCapra, F. Acridinium esters as high specific activity labels in immunoassay. Clin Chem 1983, 29(8):474-1479.

20. Pereira P, Westgard J, Encarnação P, De Sousa G. Quality management systems in the European screening laboratories in blood establishments: a view on current approaches and trends. Presented at the NRL European Workshop on Quality, IPFA/PEI 22nd International Workshop, May 19, 2014; Prague, Czech Republic.

21. Bureau International des Poids et Mesures. JCGM 100 Evaluation of measurement data - Guide to the expression of uncertainty in measurement, GUM 1995 with minor corrections. Sèvres: BIPM, 2012.

22. Pereira P, Westgard J, Encarnação P, De Sousa G. Practicability and relevance of GUM approach for computing measurement uncertainty in blood bank screening tests. Proceedings of the 2nd World Congress of Health Research, Oct 7-8, 2014; Viseu, Portugal. In: Atención Primaria, 46(Espec Cong 1):49.

23. Westgard J., Neil Carey R, Wold S. Criteria for judging precision and accuracy in method development and evaluation. Clin Chem 1974, 20(7):825-833.

24. Pereira P, Westgard J, Encarnação P, De Sousa G. The receiver operating characteristic curve and area under the curve as substitute of GUM model for determination of measurement uncertainty in blood establishments' screening tests results. Proceedings of the II International Conference of Health Research, Oliveira de Azeméis, Feb 19-21, 2014; Viseu, Portugal. In: Evidências, I Suplemento: 13.

25. Justice T. What is guard band? Solid State Technol 1997, 40(4):85–86.

26. Ellison SLR and Williams A. Use of uncertainty information in compliance assessment. Eurachem/CITAC, 2007.

27. Clinical and Laboratory Standards Institute. C24-A3 Statistical quality control for quantitative measurement procedures: principles and definitions; Approved Guideline. 3rd ed. Wayne (PA): CLSI, 2006.

28. Westgard J, Barry P, Hunt M, Groth T. A multi-rule Shewhart chart for quality control in clinical chemistry. Clin Chem 1981, 27(3):493-501.

This page intentionally left blank

Part II - Models for Determination of Uncertainty

This page intentionally left blank

Chapter 3 - Measurement Uncertainty as a Tool for Evaluating the “Grey-Zone” to Reduce the False Negatives in Immunochemical Screening of Blood Donors for Infectious Diseases

This page intentionally left blank

3.1. Abstract

The risk of misclassifying infected individuals as healthy constitutes a crucial challenge when screening blood donors by means of immunoassays. This risk is especially challenging when the numerical results are close to the clinical decision level, i.e. in the "grey-zone". The concept of using measurement uncertainty for evaluating the "grey-zone" has previously not been systematically applied in this context. This Chapter explains methods, models and empirical (top-down) approaches for the calculation of measurement uncertainty using results from a blood bank according to the internationally accepted GUM principles, focusing on uncertainty sources in the examination process. Of the different approaches available, the intralaboratory empirical approaches are emphasized since modelling (bottom-up) approaches are impracticable due to the lack of reliable model equations for immunoassays. Different methods are applied to estimate the measurement uncertainty for the Abbott Prism® HCV immunoassay. The expanded uncertainty obtained at the clinical decision level from the intralaboratory empirical approach was 36 %. The estimated uncertainty was used to set acceptance and rejection zones following the procedure set in the Eurachem guideline, emphasizing the need to minimize the occurrence of false negatives.

3.2. Introduction

Blood banks (blood establishments) collect and produce human blood components intended for transfusion. Screening immunoassays are essential in identifying blood donors infected with transmissible agents [1-4]. A positive screening immunoassay result - indicating that the donor has been infected - requires further testing by means of even more specific tests including nucleic acid amplification tests (NAT) [5].

The results of a screening immunoassay are not exact, and it is standard to define ranges of result values for which the test is considered positive, negative and inconclusive. The range of values for which the test is deemed inconclusive is usually referred to as the "grey-zone". In this paper, we address different methodologies for determining the "grey-zone" in screening immunoassays. To make it more practical,

the discussion of the different methods is done taking as example the Abbot Prism® HCV (Abbott Diagnostics, Abbott Park, IL, USA) [6].

The Abbott Prism® HCV immunoassay is an in vitro chemiluminescent immunoassay for the measurement of the concentration of antibodies to the hepatitis C virus (HCV). Persons infected with HCV produce antibodies to the virus and maintain lifelong elevated concentrations of these antibodies. The Abbott Prism® HCV test uses microparticles coated with three different recombinant HCV antigens as a solid phase which binds possible HCV antibodies present in human serum or plasma. After incubation and a washing step, the presence of immunoglobulins bound to the microparticles is measured by means of chemiluminescent anti-human IgG.

The chemiluminescent signal is proportional to the concentration of the antibodies to HCV (anti-HCV) present in the sample. Therefore, measuring the concentration of the anti-HCV is primarily to classify a blood donor as HCV-infected or not. The immunoassay is calibrated using a number of plasma samples from persons not infected by HCV (negative control, inactivated plasma recalcified and preserved nonreactive for HBsAg, HIV-1 Ag or HIV-1 NAT, anti-HCV and anti-HIV-1/HIV-2) and a number of plasma samples from individuals infected by HCV (positive control, inactivated plasma recalcified and preserved reactive for anti-HCV, nonreactive for HBsAg, HIV-1 Ag or HIV-1 NAT, and anti-HIV-1/HIV-2, minimum "cuttoff" equal to 1.25). A procedure for calculating the "cuttoff" value, using the number of emitted photons, which creates the clinical decision level, distinguishing positive and negative test results, is defined by the reagent manufacturer.

The donor's result is the ratio of the sample signal divided by the "cuttoff" value. The test results are positive if the ratio is greater than or equal to one and negative if the ratio is lower than one. Test results with a ratio close to one have a significant probability of being incorrectly classified. A "grey-zone" is normally used to indicate where there is a high frequency of false results (high b-error and/or high a-error) [7, 8]. In the author's opinion, this "grey-zone" should be evaluated using the estimated measurement uncertainty. In order to reduce the occurrence of false negative results, a decision limit can be calculated by subtracting part of the expanded measurement uncertainty from the "cuttoff" value [9].

Evaluating measurement uncertainty [10] in blood banks is not compulsory by the European Union Directives [11-14] or by the US standards [15]. Similarly, it is not required in the European Union Directives for medical laboratories [16] or in the US Clinical Laboratory Improvement Amendments (CLIA) [17]. It is, however, required for any test accredited by ISO 15189 (clause 5.5.1.4 of [18]) or by ISO/IEC 17025 (clause 5.4.6 of [19]).

Today there are mainly four different approaches to estimating measurement uncertainty [20] according to the principles laid out in the Guide to the Expression of Uncertainty in Measurement, known as the GUM [10]. This Chapter discusses the practicality and relevance of these approaches to estimating measurement uncertainty in the examination process when screening serum or plasma samples using an immunoassay and how to evaluate a "grey-zone" in order to increase the post-transfusion safety.

3.3. Materials and methods

Test results from the Portuguese Institute of Blood and Transplantation (IPST) obtained from 2010 to 2014 were used in this study. Results are from a single screening chemiluminescent immunoassay Abbott Prism® HCV immunoassay [6, 21]. In order to not underestimate the relative uncertainty, the test samples used for uncertainty estimation should be close to the "cuttoff" value.

Most of the uncertainty estimates were performed using freely available software from the Finnish Environment Institute (SYKE) MUkit, version 1.9.5.0 [22]. This software is based on the intralaboratory approach focusing on validation and quality control data laid out in a Nordtest report [23].

3.3.1. Fitness for purpose for sample results equal to or close to "cutoff" value

A critical problem in screening immunoassays occurs when the numerical result is equal to or close to the "cuttoff" value since it has a statistically significant chance of being misclassified. An interval around the "cuttoff" value, known in blood banks as the "grey-zone" [6, 7], is defined to express this uncertainty about the result. This is similar to the one-side concept referred to as the 'guard band' in other branches of metrology [24]. The interval can be determined from the uncertainty, corresponding to an acceptable level of confidence.

Table 1 displays the percentage of positive, indeterminate and negative results according to different "grey-zones" for 9805 blood samples from the IPST. The results showed a small percentage (0.58 %) of indeterminate results, even at a "grey-zone" defined as a ratio of one ±0.4. The samples with indeterminate results were tested with more specific methods. They all tested negative or indeterminate on the confirmatory immunoassay, INNO-LIA® HCV SCORE (Fujirebio Europe N.V., Ghent, Belgium) and negative for the NAT test multiplex real-time PCR Cobas® TaqScreen MPX Test, version 2.0 (Roche Diagnostics GmbH, Mannheim, Germany).

Table 1 Number and percentage of positive, indeterminate and negative results using "grey-zones" from ±5 to ±40% for 9805 blood donations from September 12 to December 21, 2014 in a blood and tissues and cells establishment of the Portuguese Institute of Blood and Transplantation.

"Grey-zone" interval	Results					
	Positive		Indeterminate		Negative	
(%)	No.	%	No.	%	No.	%
±5	41	0.42	8	0.08	9756	99.50
±10	40	0.41	12	0.12	9753	99.47
±15	39	0.40	20	0.20	9746	99.40
±20	38	0.39	23	0.24	9744	99.38
±25	37	0.38	28	0.29	9740	99.34
±30	37	0.38	34	0.35	9734	99.28
±35	34	0.35	46	0.47	9725	99.18
±40	34	0.35	57	0.58	9714	99.07

3.3.2. Analytical method of a screening immunoassay

In the Abbott Prism® HCV in vitro chemiluminescent immunoassay, the measurand is the concentration of immunoglobulin in the serum or plasma samples which bind to solid-phase particles attached recombinant antigens of the Core, NS3, NS4 and NS5 regions of the HCV genome. The signal from test samples is the detection of the light from the chemiluminescent reaction, and it is expressed in number of photons over a fixed time period. The signal is proportional to the antibody-antigen complexes, and they are corrected for measured number of photons in the dark. Positive and negative calibrators are used to establish the "cuttoff" value calculated for each analytical run measuring negative and positive calibrators according to:

$$n_{\mathrm{c}} = \overline{x}_{\mathrm{n}} + 0.55\overline{x}_{\mathrm{p}} \qquad \text{(eq. 1)}$$

where n_c is the "cuttoff" value in number of photons, $\overline{x}_n$ is the negative calibrator in number of photons expressed as the average result from the two lowest replicates out of three, and $\overline{x}_p$ is the positive calibrator in number of photons expressed as the average result from the two highest replicates out of three [6]. A correction to the number of photons is performed discounting the number of photons in the dark. This equation is set by the manufacturer, based on analysis of true negative and true positive human samples and usually approved by national agencies. A factor of 0.55 is associated. This factor is critical for the false negative rate and is set by the manufacturer.

An example of an anti-HCV test "cuttoff" determination is shown in Table 2. In the example, $\overline{x}_n$ = 616 and $\overline{x}_p$ = 32895, giving a "cuttoff" value of 20708 photons. At the end of an analytical run, a multi-parametric positive sample is used in order to control the run. This sample is produced by the manufacturer, and the acceptance criterion for anti-HCV in this case is very broad with a ratio value in the range from 1.02 to 6.00 [25].

Table 2 Example of an anti-HCV test "cuttoff" determination on equipment sub channel A, the coefficient of variation (CV) of the calibration measurements must be equal or less than 3 % (Abbott Diagnostics, Abbott Park).

	First measurement			Second measurement			Third measurement					
	Number of in the dark [a]	Raw number of photons	Corrected number of photons	Number of in the dark [a]	Raw number of photons	Corrected number of photons	Number of in the dark [a]	Raw number of photons	Corrected number of photons	Mean	SD[c]	CV
Positive calibrator	7	32140	32123	5	33662	33657	[b]	[b]	[b]	32895	1085	3 %
Negative calibrator	6	2608	2602	[b]	[b]	[b]	5	2635	2630	2616	20	1 %

[a] Mean dark counts must be ≤ 180

[b] Two out of three calibrator results are used in the calculation of the "cuttoff"; the third result is not reported by the equipment

[c] Standard deviation

The screening immunoassay ratio of the analytical method is calculated by dividing the number of photons from the sample ns by the "cuttoff" value [6]. In an analytical run, each sample produces a single measurement result. For the internal quality control (QC), a sample Accurun-1 Series 2400 was used (Seracare Life Sciences Inc., Milford, MA, USA). The results for this QC sample show a within-laboratory reproducibility standard deviation of about 10 % at ratios close to 2.5.

The calibration of the Abbott Prism® HCV method is not established from an international standard, but every lot of reagents is calibrated by measurement signals acquired from samples from persons that have not been infected by HCV (negative) and persons infected by HCV (positive). Several types of bias may occur in laboratories [26-28]. The calibration procedures used for the present method mean that the within-laboratory uncertainty component, day to day, caused by bias over time will be random. We have not been able to properly assess a true between-laboratory bias since certified reference materials (CRM) or a reference laboratory was unavailable. However, parallel measurements of the same control sample over a 2-month period in two different laboratories, using the same method, showed a difference in the mean values of 6 % at a ratio of 2.4.

3.3.3. The approaches to measurement uncertainty

In screening immunoassays, measurement uncertainty provides information on the level of confidence that can be placed on the positive or negative result when comparing a numerical result with a "cuttoff" value. Measurement uncertainty is the quantitative expression of the doubt associated with the result. The expanded uncertainty *U* provides an interval within which the value of the measurand is believed to lie with a specified level of confidence. U is obtained by multiplying the standard combined uncertainty $u_c(y)$ of a value y by a coverage factor *k*:

$$U = k \cdot u_c \qquad \text{(eq. 2)}$$

The choice of the factor k is based on the level of confidence desired. For an approximate level of confidence of 95 %, k is usually set to 2, and for a confidence higher than 99 %, k is usually set to 3 for more than 20 degrees of freedom for the combined uncertainty.

The result is often presented in the format, $x \pm U$ *for* an observed value x. For example, suppose that in a blood test, the screening immunoassay gives a ratio of 1.1 ± 0.2 using an expanded uncertainty $k = 2$, corresponding to the interval 0.9-1.3, with a defined "cuttoff" at a ratio of one. This result is interpreted as: $x \pm U$ contains the "cuttoff" value, and thus the result is in an indeterminate zone where we cannot declare it to be positive or negative.

More than 20 years after the publication of the first edition of the GUM, this document is still acknowledged as the master document on estimating measurement uncertainty throughout the testing community [10]. In the chemistry sector, the Eurachem Uncertainty Guide, which follows the GUM principles, is widely accepted [29].

The basis for any valid uncertainty evaluation is:

- A clear definition of the measurand, i.e. the quantity to be measured,
- A comprehensive specification of the measurement procedure and the measurement objects, and,
- A comprehensive analysis of the effects impacting the measurement results. From a laboratory view, the main effects are intermediate precision (within-laboratory precision) and any residual measurement bias. For a detailed discussion on treatment of the bias term, see Magnusson and Ellison [28].

Following the GUM principles, according to a Eurolab report [20], there are four main approaches to estimate measurement uncertainty:

1. Modelling
2. Single-laboratory validation (including QC)
3. Interlaboratory comparisons
4. External quality assessment (EQA) [proficiency testing (PT)]

Approach 1 is mainly based on a model equation, whilst approaches 2, 3 and 4 are mainly based on experimental data. In many cases, a combination of several approaches is applied in order to estimate measurement uncertainty. For each approach, the combined standard uncertainty u_c is calculated and a factor k is chosen dependent on the confidence level.

The modelling approach for the evaluation of uncertainty is described in chapter 8 of the GUM. This procedure is based on a model, often expressed in the form of an equation to account for the interrelation of the input quantities that influence the measurand combined with additional experimental data. A correction is included in the model to account for any recognized systematic effects. The application of the law of propagation of uncertainty [10] or Monte Carlo simulation [30] enables the evaluation of the result combined standard uncertainty u_c. In the case of a screening immunoassay test, any model equation to describe the measurement is in detail, to our knowledge, not known, and therefore the modelling approach cannot be applied.

In the single-laboratory validation and QC approach (intralaboratory), the major sources of variability can often be assessed by an in-house validation study combined with ongoing internal QC by means of repeated measurements of stable control samples. The use of CRM and/or comparison with reference methods can help to evaluate the component of uncertainty related to possible bias. The outline of this approach is described in detail in Nordtest 537 [23] and ISO 11352 [31] for environmental analysis, but the approach is generic and can be applied to many different analyses.

In the interlaboratory validation approach, the major sources of variability can often be assessed by interlaboratory studies and evaluated according to ISO 5725 [32]. This approach to estimating uncertainty is fully described in ISO 21748 [33].

In the EQA approach, data from proficiency tests are used. Such tests are intended to check periodically the overall performance of a laboratory. The laboratory's results from its participation in proficiency testing can accordingly be used to check the evaluated uncertainty, since that uncertainty should be comparable with the spread of results obtained by that laboratory over a number or proficiency test rounds. Results from proficiency testing can also be used to evaluate the uncertainty. If the same method was used by all the participants in the EQA scheme, then the standard deviation calculated from the individual results could be considered as a preliminary evaluation of the combined standard uncertainty [20].

It should be noted that it still is a challenge to reliably estimate measurement uncertainty. Thompson and Ellison (2011) argued that usually the estimated measurement uncertainty is significantly lower than the standard deviation under reproducibility conditions which indicates an underestimation of the uncertainty. The uncertainty not recognized was referred to as “dark uncertainty” [34].

3.3.4. Estimation of measurement uncertainty

1. Single-laboratory validation approach for estimating uncertainty

This approach is primarily intended for the validation of quantitative tests, considering precision and bias components. In this case, the quantitative result is the ratio obtained in the immunoassay and is used to calculate the uncertainty for this screening test. The within-laboratory reproducibility uncertainty s_{Rw} is calculated by pooling the repeatability standard deviation s_r arising from replicate measurements of human samples having a ratio from 0.5 to 1.5 and the intermediate standard deviation, s_I, from between-runs as in Eq. 3 [23]:

$$s_{Rw} = \sqrt{s_r^2 + s_I^2} \quad \text{(eq. 3)}$$

Equation 4 is used to determine the bias uncertainty u_b, if only one CRM is used. The bias, b, is the mean deviation of replicate measurement results from the corresponding reference value, s_b is the standard deviation of the bias measurements, $u(c_{ref})$ is standard uncertainty of the certified reference value, and m is the number of replicate measurements:

$$u_b = \sqrt{b^2 + (s_b/\sqrt{m}) + u(c_{\mathrm{ref}})^2} \quad \text{(eq. 4)}$$

To obtain the combined standard uncertainty of a value y, the within-laboratory reproducibility uncertainty s_{Rw} and the bias uncertainty u_b are combined as in Eq. 5 [20]:

$$u_c(y) = \sqrt{s_{\mathrm{Rw}}^2 + u_b^2} \quad \text{(eq. 5)}$$

This section considers the within-laboratory reproducibility standard deviation according to two different methods, (a) and (b):

(a) Validation protocol: it is intended for validating the precision of numerical quantity (quantitative) tests described in Clinical Laboratory Standards Institute (CLSI) EP15-A3 protocol and can be used to evaluate the precision of a screening immunoassay [35]. Using this method, data should be collected for a sample with a concentration close to the "cuttoff" concentration. A series of five analytical runs with three replicates per run is usually employed.

(b) Using data from between-run variation. For the case of the Abbott Prism® HCV immunoassay that is being used as an example in this paper, data regarding the "cuttoff" number of emitted photons from 299 runs using 9 reagent batches, from 18 April 2010 to 10 November 2011, were used. Data for laboratory bias evaluation were not available since there is no CRM or reference laboratory. We used the mean difference between two laboratories of the IPST using the same analytical equipment as an estimator of the bias. The bias, due to the use of different reagent batches, is already included in the within-laboratory reproducibility standard deviation.

2. Interlaboratory comparison approach

The approach requires the determination of the between laboratory reproducibility standard deviation s_R using the results in an interlaboratory trial according to ISO 5725. In a standardized method, these precision data are normally given in an appendix to the procedure. However, the precision data for the Abbott Prism® immunoassay are not yet available.

3. External quality assessment

The purpose of an EQA program is to verify periodically the overall performance of a laboratory against a group of laboratories [36]. From the results obtained in all laboratories, the reproducibility standard deviation can be calculated and used as an estimate of standard combined uncertainty. In this approach, both the day-to-day bias and the laboratory bias are included. This approach is not applicable to programs where only qualitative results are available.

3.3.5. How to calculate the decision limit

We suggest adopting the approach given in the Eurachem/CITAC Guide [9] for the use of uncertainty in compliance assessment whilst focusing on reduction in false negatives in order to reduce the risk of post-

transfusion infection. Two zones need to be set up, (1) a “rejection zone” where the results are treated as positive and (2) an “acceptance zone” where results are assured to be from true negative samples. The intersection between these two zones is called the decision limit. The decision limit is calculated by subtracting $1.65u_c$ from the “cuttoff” ratio value, assuming that the results are normally distributed and that the number of degrees of freedom is large enough to warrant using the value 1.65 from the one-tailed *t* value at 95 % confidence. For an in depth discussion of the use of uncertainty in compliance assessment, further information can be found in [9].

3.4. Results

In most instrumental methods, measurement results have a relative uncertainty and a relative standard deviation under repeatability and reproducibility conditions that decrease with the increase in the analyte level (section E.4 in [29]). That was also the case for the Abbott Prism® HCV immunoassay where the relative uncertainty decreased with the increase in the concentration. Table 3 summarizes the estimates of the uncertainty from the intralaboratory and EQA approaches using available experimental data at different ratios. As can be seen from the results in this table, the relative standard deviation decreases as the ratio increases and the highest relative standard deviation is observed in the range for test samples close to a ratio of one - the repeatability results based on duplicate measurement with ratio in the range from 0.5 to 1.5.

Table 3 Estimated measurement uncertainty values for the Abbott Prism HCV method using data from intralaboratory and EQA approaches.

		Within-lab reproducibility u_{Rw}							Bias uncertainty	Combined uncertainty	Expanded uncertainty
Approach	Method	Ratio	s_r (method)	s_r	s_I (method)	s_I	s_{Rw} (Eq. 3)	Bias method	u_b	u_c	U
Intralaboratory	EP 15 validation data[b]	2.2 - 4.2	NA	NA	NA	NA	8.5%	Control sample measured at two laboratories [b]	6.0%	10.4%	21%
	Validation and QC data	0.5 - 2.8	Replicates from ratio 0.5 to 1.5	14.4%	Between-run "cuttoff"[c]	8.5 %	16.5%	Control sample measured at two laboratories[b]	6.0%	17.8%	36%
EQA[a]		6.2 - 10.1	NA	NA	NA	NA	NA	NA	NA	13.9%	28%

[a] For this approach the obtained reproducibility standard deviation is set as the combined uncertainty, sample UK NEQAS no. 9316, average ratio equal to 2.67 (results from 9 laboratories)
[b] Sample Accurun-1 Series 2400 batch no. 10017751 (Seracare Life Sciences Inc., Milford, MA, USA); average ratio equal to 3.70 (15 replicate measurements per run)
[c] Average ratio equal to one (299 results per run)
NA, not applicable

3.5. Discussion

3.5.1. Estimation of measurement uncertainty

In a manufacturer's precision study, using data from five runs over 5 days under intralaboratory conditions (within laboratory reproducibility), a s_{Rw} from 5.7 % (average ratio 3.17) to 8.6 % (average ratio 0.17) was obtained [6]. A similar precision estimate (s_{Rw} = 8.5 %) was obtained by us using EP15 validation data. On the other hand, the validation and QC data intralaboratory approach using replicated results and between-run "cuttoff" data estimated s_{Rw} in 16.5 %. In this case, the major uncertainty component was the standard deviation of the replicate results. This is due to the fact that we have selected results only where the ratio is close to one. At higher ratios, the relative standard deviation is lower as can be seen from the EP15 results with an average ratio equal to 3.70.

Comparing the use of the EP15 precision data with the use of the between-run precision data in the second intralaboratory approach, the reliability of the between-run data is higher, since the number of degrees of freedom of this estimate is much higher. The results were derived from close to 300 measurement results obtained over a longer time period (close to 4 years), whereas EP15 data are based on only 15 measurement results over a time period of 5 days (16–20 December 2014).

The EQA data from one round with nine results all using the Abbott Prism® provided an estimate for the relative expanded uncertainty [32] of 28 %, whilst the validation and QC data intralaboratory approach estimated uncertainty in 36 %. The heterogeneity of the group participants is the principal cause for a higher measurement uncertainty. In this case, a sample with an average ratio equal to 8.2 was used. If a sample with lower ratio, closer to one, had been used, the estimated uncertainty would have been even higher.

Allegedly, the use of long-term internal QC data could be an alternative to the between-run imprecision using "cuttoff" results since it covers the whole analytical process. However, the use of a QC material with an average ratio of one makes it a secondary option. For example, using internal QC data from 23 March 2010 to 11 June 2011 testing sample Seracare Accurun-1 Series 2400 batch no. 116406 (Seracare Life Sciences Inc., Milford, MA, USA) with an average ratio of 2.42 (based on 293 results per run), an s_I equal to 11.1 % was obtained. Because of the higher average ratio, the method was deprecated in favor of between-run "cuttoff" data.

It is recommended the use of the intralaboratory approach using EP15 data as an initial uncertainty estimation in a brand new test. When long-term data are available, it is recommended to apply the use of the intralaboratory approach using repeatability from ratio 0.5 to 1.5 and between-run "cuttoff" precision.

3.5.2 The "grey-zone" and the decision limit

With a combined uncertainty uc of 18 %, the decision limit calculated according to the Eurachem Guide [9] will be 30 % lower than the ratio of 1.00 when focusing on low false-negative rate. Thus, results above or equal to a ratio of 0.70 will be in the 'rejection zone', and the corresponding blood components are declared not compliant and are rejected. Results that are less than a ratio of 0.70 are in the 'acceptance zone', and the corresponding blood components are in compliance with the intended use of results and are accepted. Using a decision limit of a ratio of 0.70 in a blood bank has a low impact on the blood bank's production (waste of blood bank budget) since the number of test samples in this zone with a ratio from 0.70 to 1.00 is only 0.19 % of 9805 samples. The approach of calculating a decision limit focusing on reducing the false negatives can be applied to all screening immunoassays as well as to other qualitative

tests. A limitation to the present study is the possibility that the manufacturer “cuttoff” equation already considers the manufacturer decision limit. If that is the case, the decision limit of the manufacturer could be used directly, providing the uncertainty obtained, since conditions in laboratories are similar to the conditions in which the manufacturer developed the equation for setting the “cuttoff” value. However, the conditions can indeed vary from country to country and need to be taken into account.

3.6. Conclusions

The uncertainty estimation approach is generic and can be applied to any other immunoassay, independently of the measurement method. In order to reduce the false negatives due to analytical sources, the “grey-zone” around the ‘cutoff’ value needs to be reliably determined. We have estimated this “grey-zone” using guidelines for measurement uncertainty and for compliance assessment. Of the four different main approaches to estimating uncertainty, three approaches have been investigated: the two intralaboratory approaches and one EQA approach. The other approaches to estimating uncertainty are at present not feasible. The three approaches used gave a relative expanded uncertainty between 21 % and 36 %, where the higher uncertainty estimate at a ratio close to one using long-term data for repeatability with ratio from 0.5 to 1.5 and between-run “cuttoff” precision is the most reliable. This approach using a “grey-zone” based on measurement uncertainty to reduce false negatives is generic and can be applied to any similar and similarly calibrated immunoassays.

References

1. White G. Risk management in blood transfusion services. Vox Sang 1998, 74(S2):105-109.

2. Soldan K, Barbara J, Ramsay M, Hall A. Estimation of the risk of hepatitis B virus, hepatitis C virus and human immunodeficiency virus infectious donations entering the blood supply in England, 1993-2001. Vox Sang 2003, 84(4):274-286.

3. Pereira P, Westgard J, Encarnação P, Seghatchian J, Sousa G. Quality management in the European screening laboratories in blood establishments: a view on current approaches and trends. Transfus Apher Sci 2015, 52(2):245-251.

4. Pereira P, Westgard J, Encarnação P, Seghatchian J, Sousa G. The role of uncertainty in results of screening immunoassays in blood establishments. Transfus Apher Sci 2015, 52(2):252-255.

5. Williams E, Jarreau P, Zitzmann M, Pitocco C. Transfusion-transmitted diseases. In: Harmening D (editor). Modern blood banking & transfusion practices. 6th ed. Philadelphia: F.A. Davis, 2012; p. 403-426.

6. Abbott Diagnostics Division. Abbott Prism HCV. Code: 84-5342/R9. Wiesbaden: Abbott Laboratories, 2008.

7. Coste J, Jourdain P, Pouchot J. A gray zone assigned to inconclusive results of quantitative diagnostic tests: Application to the use of brain natriuretic peptide for diagnosis of heart failure in acute dyspneic patients. Clin Chem 2006, 52(12):2229-2235.

8. Kisner HJ. The gray zone. Clin Lab Manage Rev 1998, 12(4):277-280.

9. Ellison SLR and Williams A. Use of uncertainty information in compliance assessment. Eurachem/CITAC, 2007.

10. Bureau International des Poids et Mesures. Evaluation of measurement data - Guide to the expression of uncertainty in measurement. JCGM 100:2008, GUM 1995 with minor corrections. Sèvres: BIPM, 2008. Retrieved from http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf. Accessed: December 22, 2015.

11. Commission Directive 2002/98/EC of the European Parliament and of the Council as regards setting standards of quality and safety for the collection, testing, processing, storage and distribution of human blood and blood components and amending Directive 2001/83/EC. Official Journal of the European Union L33/30-40.

12. Commission Directive 2004/33/EC Implementing Directive 2002/98/EC of the European Parliament and of the Council as regards certain technical requirements for blood and blood components. Official Journal of the European Union L91/25-39.

13. Commission Directive 2005/61/EC Implementing Directive 2002/98/EC of the European Parliament and of the Council as regards traceability requirements and notification of serious adverse reactions and events. Official Journal of the European Union L256/32-40.

14. Commission Directive 2005/62/EC Implementing Directive 2002/98/EC of the European Parliament and of the Council as regards Community standards and specifications relating to a quality system for blood establishments. Official Journal of the European Union L256/41-48.

15. American Association of Blood Banks. Standards for blood banks and transfusion services. 29th ed. Bethesda (MD): AABB, 2014.

16. Commission Directive 98/79/EC of the European Parliament and of the Council on in vitro diagnostic medical devices. Official Journal of the European Union L331/1-3712.

17. Clinical Laboratory Improvement Amendments. CLIA law & regulations, 2014. Available at: http://wwwn.cdc.gov/clia/Regulatory/default.aspx. Accessed: December 4, 2015.

18. International Organization for Standardization. ISO 15189 Medical laboratories - Requirements for quality and competence. 2nd ed. Geneva: ISO, 2012.

19. International Organization for Standardization/International Electrotechnical Commission. ISO/IEC 17025 General requirements for the competence of testing and calibration laboratories. 2nd ed. Geneva: ISO, 2005.

20. European Federation of National Associations of Measurement, Testing and Analytical Laboratories. Technical Report No. 1/2007 Measurement uncertainty revisited: Alternative approaches to uncertainty evaluation. EuroLab, 2007.

21. Khalil OS, Zurek TF, Tryba J, Hanna CF, Hollar R, Pepe C, et al. Abbott prism: a multichannel heterogeneous chemiluminescence immunoassay analyzer. Clin Chem 1991, 37(9):1540-1547.

22. SYKE. MUkit - Measurement Uncertainty Kit Finnish Environmental Institute, 2014. Retrieved from: http://www.syke.fi/en-us/Services/Calibration_services_and_contract_laboratory/MUkit__Measurement_Uncertainty_Kit Accessed: December 22, 2015.

23. Magnusson B, Näykk T, Hovind H, Krysell M. NordTest NT TR 537 Handbook for Calculation of Measurement Uncertainty in Environmental Laboratories. 3.1th ed. Oslo Nordic Innovation, 2011.

24. Justice T. What is guard band? Solid State Technol 1997, 40(4):85-86.

25. Abbott Diagnostics Division. Abbott Prism positive run control kit HCV. Code: 34-3827/R7. Wiesbaden: Abbott Laboratories, 2005.

26. Hovind H, Magnusson B, Krysell M, Lund U, Mäkinen I. NordTest NT TR 569 Internal Quality Control - Handbook for Chemical laboratories (Trollboken - Troll book). 4th ed. Oslo: Nordic Innovation, 2011.

27. Theodorsson E, Magnusson B, Leito I. Bias in clinical chemistry. Bioanalysis 2014, 6(21):2855-2875.

28. Magnusson B, Ellison SL. Treatment of uncorrected measurement bias in uncertainty estimation for chemical measurements. Anal Bioanal Chem 2008;390(1):201-213.

29. Ellison SLR, Williams A. Quantifying uncertainty in analytical measurement. 3rd ed. Eurachem/CITAC, 2012

30. Bureau International des Poids et Mesures. JCGM 101 Evaluation of measurement data - Supplement 1 to the “Guide to the expression of uncertainty in measurement” - Propagation of distributions using a Monte Carlo method. Sèvres: BIPM, 2008.

31. International Organization for Standardization. ISO 11352 Water quality - Estimation of measurement uncertainty based on validation and quality control data. Geneva: ISO, 2012.

32. International Organization for Standardization. ISO 5725-2 Accuracy (trueness and precision) of measurement methods and results - Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method. Geneva: ISO, 1994.

33. International Organization for Standardization. ISO 21748 Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation. Geneva: ISO, 2010.

34. Thompson M, Ellison S. "Dark uncertainty." Accred Qual Assur 2011, 16(10): 483-487.

35. Clinical and Laboratory Standards Institute. EP15-A3 User verification of precision and estimation of bias; Approved guideline. 3rd ed. Wayne (PA): CLSI, 2014.

36. Walker S, Dimech W, Kiely P, Smeh K, Francis B, Karakaltsas M, Dax E. An international quality control programme for PRISM chemiluminescent immunoassays in blood service and blood product laboratories. Vox Sang 2009, 97(4):309-316.

Chapter 4 - Complementary Approaches to Measurement Uncertainty: Total Analytical Error, Seroconversion Window Period and Delta Value

This page intentionally left blank

4.1. Abstract

The evaluation of measurement uncertainty is not required by the European Union regulation for blood establishments' laboratory tests. However, it is required for tests accredited by ISO 15189. ISO recommends GUM models for determination of measurement uncertainty, but their application is not feasible when dealing with ordinal value measurements, such as what happens with screening test binary results. The impact of uncertain results in blood establishment's budget is usually not estimated. This Chapter reviews, discusses and proposes complementary models to measurement uncertainty of screening test results. The total analytical error model combines variance and bias, and it is considered by some as an alternative to measurement uncertainty determination. The seroconversion window period determines the seronegative period of an infected individual and cannot be combined in measurement uncertainty. The delta-value of infected and healthy individuals' samples allows ranking two or more tests according to the probability of generate indeterminate results, allowing a simple evaluation of risk of waste on budget.

4.2. Introduction

As it was discussed in Chapter 1, ISO 15189 guideline (2012) requires the estimation of measurement uncertainty as well as the specification of the performance requirements (quality goal) for measurement uncertainty and the regular review of the its estimation [1]. Measurement uncertainty determination is optional on blood establishment laboratories field, since it is not mandatory by the European Union Directives [2-5] or by US Clinical Laboratory Improvement Amendments (CLIA) requirements integrated in the American Association of Blood accreditation program [6]. ISO 9001 has been widely applied in European blood establishments [7] and the introduction of "risk based thinking" could be supported in blood establishment laboratories by determining measurement uncertainty (entry 3.09 of [8]).

ISO recommends determination of measurement uncertainty by following the Guide to the Expression of Uncertainty in Measurement (GUM) principles [9]. Dimech et al. successfully applied a GUM modeling approach [10] and an interlaboratory empirical approach [11] to a screening test. However, the GUM

concept is focused on numerical quantity values (entry 1.20 of [12]), rather than ordinal quantity (entry 1.26 of [12]) tests providing binary results (i.e., positive/negative). For an in depth discussion of GUM measurement uncertainty approach, further information can be found elsewhere [9]. Ordinal quantity is defined in the International Vocabulary of Metrology (VIM) as the "quantity, defined by a conventional measurement procedure, for which a total ordering relation can be established, according to magnitude, with other quantities of the same kind, but for which no algebraic operations among those quantities exist", and it is recognized in medical laboratories as "qualitative" quantity.

Chapter 2 discussed causes to uncertain results not including uniquely measurement uncertainty, but also others sources and some are unmeasurable according to measurement uncertainty concept. One of the most easily recognized statistical tools in medical laboratories is the total analytical error, used principally in method validation of clinical chemistry. Total analytical error (TAE) has been considered by some as an alternative concept to measurement uncertainty [13]. Seroconversion window period is easily associated with seronegative period in infected individuals and it is usually determined in the evaluation of screening tests. The period is cannot be combined in standard uncertainty, for what its result is complementary to measurement uncertainty. The effect of uncertain results in blood establishments' budget (waste) is not frequently evaluated. The delta-value is a simple model that allows to rank a test in a series of tests according to the chance to produce indeterminate results. These three models are discussed in this Chapter.

Several other authors proposed other complementary approaches to measurement uncertainty. Pulido et al. proposed for general chemistry the use of statistical intervals for the estimation of uncertainty of binary results to evaluate compliance with supervisory limits [14] and the performance curve to determine the "cutoff" concentration with an uncertainty interval [15]. It was proposed additional alternative (to GUM) concepts for diagnostic sensitivity and specificity, such as the Bayesian theorem on conditional probability [14-17] (see Chapter 5).

The data calculations for discussed models have been performed using standard spreadsheet software (Microsoft Excel® 2016).

4.3. Methods and materials

4.3.1. Determination of "grey-zone" using total analytical error

Analytical random and systematic error components at the "cutoff" concentration result are major causes of lack of stability of measurement that influence the trueness of binary results (see Chapter 3). The analytical random error could be estimated through precision in reproducibility condition, i.e., the "condition of measurement, out of a set of conditions that includes different locations, operators, measuring systems, and replicate measurements on the same or similar objects" (entry 2.24 of [12]).

Despite ISO claims "the deviation from the true value is composed of random and systematic errors. The two kinds of errors, assumed to be always distinguishable, have to be treated differently. No rule can be derived on how they combine to form the total error of any given measurement result, usually taken as the estimate" [12], TAE sums the random and systematic error in a single result remaining as an alternative concept to GUM model. It is a concept linked to the Error Approach (also recognized as Traditional Approach or True Value Approach) instead to the current Uncertainty Approach contrary to what happens with measurement uncertainty. TAE is well recognized by reagents' manufacturers [18] and the medical laboratory field [19]. It is also required by national agencies such as FDA which provides manufacturers and FDA staff with guidance on "Waiver applications for manufacturers of in vitro diagnostic devices" endorsing the use of TAE for demonstrating the performance of US waived tests [20]. Despite it was initially applied in clinical chemistry quantitative tests, it is also applied in screening immunoassays principally by reagents' manufacturers to determine the "grey-zone" [18].

The TAE model was introduced to clinical chemistry by Westgard, Carey and Wold (1974) [21]. The theory was a modification of total analytic error claimed by Eisenhart [22]. Currently it is defined by FDA as "the combination of errors from all sources, both systematic and random. It is often expressed in terms of an interval that contains a specified proportion (e.g., 95 %) of the distribution of observed differences between the waiver method and comparative method values (…). Sometimes relative differences (…) are used instead of differences." [20]. Another definition comes from CLSI EP21-A (2003) which defines TAE as the "result of a measurement minus a true value of the measurand" containing random and systematic

effects (entry 3 of [19]). Its principle is to measure assay errors arising from data collection experiment. The mathematical model for TAE measurement in laboratory medicine:

$$TAE = b_{\text{totl}} + \Delta SE_{\text{cont}} s_{\text{meas}} + z s_{\text{meas}} \quad \text{(eq. 1)}$$

where b_{totl} is the absolute value of the algebraic sum of the stable bias observed with the matrix bias, ΔSE_{cont} is the "sensitivity" of the QC procedure for detecting systematic error or changes in inaccuracy, s_{meas} is the stable measurement standard deviation or analytical imprecision, and z is related to the confidence level selected. A simpler formula includes only total bias and total imprecision:

$$TAE = b + k \cdot s_{\text{Rw}} \quad \text{(eq. 2)}$$

where b is the absolute value of bias, s_{Rw} is the within-laboratory reproducibility and k is the coverage factor based on the level of confidence desired.

The TAE measurement is performed combining the estimate of bias from a comparison of tests' study and the measures of imprecision from a replication study under reproducibility conditions (entry 2.24 of [12]). Thus, it expresses the combined effect of all the error components arising within the measurement procedure in a single value.

Bias is the "estimate of a systematic measurement error" (entry 2.18 of [12]). For example, the difference between the average of results in repeatability conditions and the assigned quantity value of a certified reference material (CRM) or of material tested in a reference laboratory. The types of bias in immunoassays are classified as:

- Proportional bias modeled by $b_{\text{p}} = (n_{\text{s}} + (m \cdot n_{\text{s}})) / (n_{\text{c}} + (m \cdot n_{\text{c}}))$ where n_{s} is the sample value, n_{c} is the "cutoff" value, and m is the constant slope with y intercept equal to zero, for example, proportional bias effect on samples and calibrators used for "cutoff" determination (note: bias is zero since $(n_{\text{s}} + (b \cdot n_{\text{s}})) / (n_{\text{c}} + (b \cdot n_{\text{c}})) = (1 + b) / (1 + b) \cdot n_{\text{s}} / n_{\text{c}} = n_{\text{s}} / n_{\text{c}}$); and
- (b) Fixed bias modeled by $b_{\text{f}} = (n_{\text{s}} + (a + b \cdot n_{\text{s}})) / (n_{\text{c}} + (a + b \cdot n_{\text{c}}))$ where b is the constant slope and a is the y intercept where $(a + b \cdot n_{\text{s}}) / (a + b \cdot n_{\text{c}}) \neq 0$, for example, different bias effect on sample and calibrators such as sample bias due to nonconforming sample centrifuge and "cutoff" bias due to nonconforming reagent storage during transport from manufacturer to laboratory [23].

The statistical process for estimating the relationships among fixed bias variables is generally impracticable in the medical laboratory field. They usually measurable at the manufacturer field.

Precision under repeatability conditions s_r (within-run precision) (entry 2.20 of [12]) and the intermediate standard deviation under reproducibility conditions s_I (within-laboratory precision) are combined to estimate the S_{Rw}. The model for the estimation of the within-laboratory reproducibility is similar to the S_{Rw} component in combined standard uncertainty u of a value y (refer to the equation 5 in Chapter 3) and the product of its result by a factor k is similar to the "expanded uncertainty" U (see Chapter 3). Generally, when the conditions of the Central Limit Theorem are met, a value for the coverage factor is taken from a one or two-tailed *t*-Student distribution with effective degrees of freedom v_{eff} for $v_{eff} < 5$ (please refer to [24] for the computation of v_{eff}). When $v_{eff} > 5$, k is assumed to be 2, corresponding roughly to the 95 % confidence interval (95 % CI). In laboratories statistics is also used $k = 1.65$ or $k = 1.96$, respectively for a one sided or two sided estimate.

Since there is no available a CRM or a reference laboratory, measurement bias was determined from the mean difference between two laboratories of the IPST using a common test. Pereira et al. (2014) demonstrated in a case that the TAE calculation is unfeasible when bias is unmeasurable, being equal to the within-laboratory reproducibility [25]. s_{RW} could be determined according to the ISO 5725 [26]. See Chapter 3 for details about its application.

The estimation can be periodically reviewed to verify compliance such as when it is considered another claimed indeterminate interval or when critical test error components changed.

4.3.2. Seroconversion window period

According to [27], "the window period for a test designed to detect a specific disease (particularly an infectious disease) is the time between first infection and when the test can reliably detect that infection". Typically, the window period is described as the number of days between the day of infection and the first positive result [28]. The term is also synonymous of "seroconversion sensitivity" [29]. For the screening immunoassays used in blood establishment virology laboratories, the seroconversion window period is the expression of the time taken for seroconversion. That is the reason for the examination of blood

donors before the donation as part of minimum eligibility criterion order to suspend or to exclude candidates with high risk to be on seroconversion period [2,30].

The performance requirement should be the highest classification in tests' rank for seroconversion, commonly available from commercial panels' inserts. A minimum seroconversion window period is also related to a test generation. Theoretically, the window period could consider the day of the first indeterminate result sample (i.e., in this case a false negative result) instead of the day of the first positive result sample, since the blood donations of indeterminate and positive results samples donors shall be rejected according to a screening algorithm. Depending on the screening panel, this may result in a shorter interval or not. For an in depth discussion of seroconversion window period, further information can be found elsewhere [31].

4.3.3. Delta-value

The delta-value is appropriate to compare test methods and rank them according to their capabilities to generate numerical results close to the "cutoff", i.e., with a significant probability to be on the indeterminate interval. It is defined "as the distance of the mean of the distribution from the "cutoff" in standard deviation units" [32]. The test with the least chance of producing indeterminate results should be preferred, since this choice reduces the chance of lack of sustainability of blood components production, i.e., decreases the probability of blood establishment budget waste.

Consider a sample of infected individuals and another of healthy individuals. Let s_{rd} and $\overline{x}_{\mathrm{d}}$ be the standard deviation and the infected individuals' mean of $\log_{10}(n_s / n_c)$, i.e., of the test results for the infected individuals expressed as the logarithm to base 10 of the ratio n_s / n_c. s_{rnd} and $\overline{x}_{\mathrm{nd}}$ refer to the same quantities for the healthy individuals' sample. The delta measurements are given by delta-plus

$$\delta+ = \overline{x}_{\mathrm{d}}(\log_{10} n_s / n_c) / s_{\mathrm{d}}(\log_{10} n_s / n_c) \quad \text{(eq. 3)}$$

where result is ≥ 1 and delta-minus

$$\delta- = \overline{x}_{nd}(\log_{10} n_s / n_c) / s_{\mathrm{nd}}(\log_{10} n_s / n_c) \quad \text{(eq. 4)}$$

where the output is < 1. The test from a series with delta-plus closest to one has the highest probability of indeterminate results. On the other hand, the test from a series with delta-minus closest to one has

also the highest probability of indeterminate results. The performance goal is to select the test with the lowest chance to generate results close to the "cutoff" value. The most uncertain test from a series of tests using same data for evaluation will have the highest chance to produce indeterminate results. For an in depth discussion of delta-value in blood establishments, further information can be found elsewhere [32].

4.4. Results and discussion

4.4.1 Total analytical error

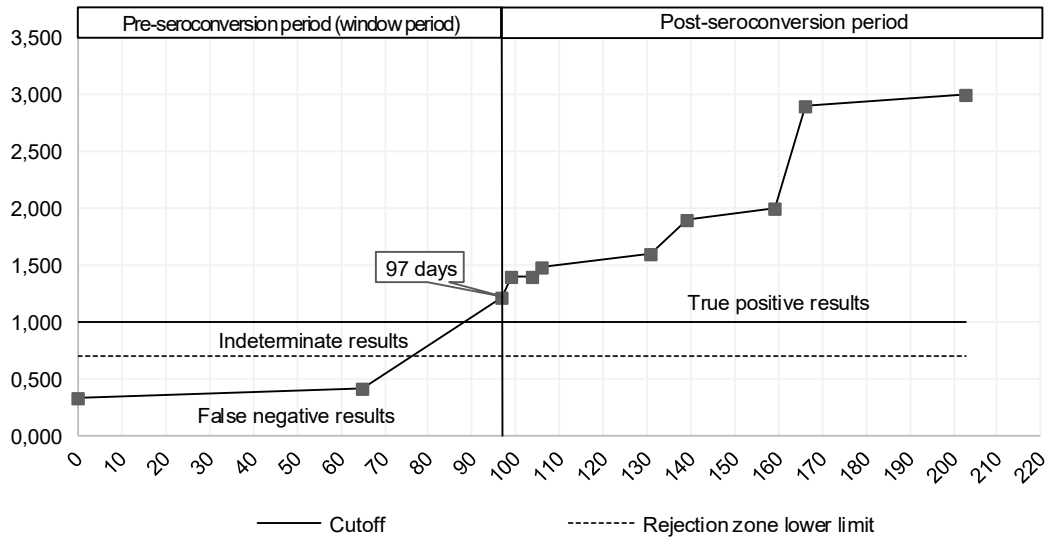
To exemplify the three models, a single screening anti-hepatitis C virus (HCV) immunoassay was used (Abbott Prism® HCV) (See Chapter 3 for details about the test). The within-laboratory reproducibility and bias were determined using an internal quality control sample Accurun-1 Series 2400 (Seracare Life Sciences Inc., Milford, MA, USA), s_{Rw} = 16.5 % and, b = 6.0 %. See Table 3 in Chapter 3 for details about sampling and calculus of the within-laboratory reproducibility with the "Validation and QC data". Consequently, TAE = 6.0 % + 2 · 16.5 % = 39 %. Comparing the TAE result with expanded uncertainty, 36 %, the TAE is overestimated and otherwise will be underestimated. It was adapted to TAE the Eurachem/Citac Guide to harmonize the calculus of decision limit between TAE and measurement uncertainty [33]. Therefore, the sum of within-laboratory reproducibility with k = 1 and bias was considered to be similar to the standard uncertainty. Consequently, the decision limit will be 37 % lower than ratio of 1.00 when focusing on low false-negative rate. The results above or equal to a ratio of 0.63 will be in the "rejection zone".

The TAE use in medical laboratories in the present is not consensual such as evidenced in the first strategic European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) conference (2014) which defined five working groups with the "TFG 3 group" that will write a manuscript dealing with "measurement total error" to recommend how to use the TAE concept and to discuss if it should be withdrawn [34]. The TAE remains in clinical chemistry practice contrary to the measurement uncertainty which is not systematically used, despite TAE not be part of current metrology theory.

4.4.2 Window period

It was tested an HCV seroconversion panel to measure the window period using not only the "cutoff" but also the decision limit 30 % lower than ratio of 1.00. A period of 97 days was necessary until the test was able to reliably detect the infection of infected individuals' panel (see Figure 7). The panel insert did not feature the test under evaluation but 11 other tests featuring nine with first positive results from the 97th day, one from the 104th day, and another from the 131st day. Considering the "cutoff" or the indeterminate interval, the infected individual's seroconversion panel tested in the assay under evaluation, obtained a window period rated at the shorter period level. The seroconversion window period aims at determining the most critical measurable bias (biological bias) component in blood establishment virology immunoassays, since it represents the major component of the risk of post-transfusion infection [31]. It is also a diagnostic accuracy test despite not generally classified as such to distinguish from Bayesian models (see Chapter 4).

Figure 1 Window period in a screening immunoassay for the PHV901 panel (SeraCare Life Sciences, Inc., Milford, MA) featuring 11 genotype 1a samples from a blood or plasmapheresis donor who seroconverted over the course of their donation history.



The definition of the seroconversion window period makes its determination dependent of the seroconversion period from a single infected individual when only a panel is tested. The days between first bleed and the first indeterminate result cannot be inferred to the infected individuals' population. Accordingly, a definition is proposed: "the window period for a test designed to detect a specific disease (particularly an infectious disease) is the time between the first day of infection and the day when the test

result cannot reliably rule out the possibility of infection (due to indeterminate results)" [25]. Window period is a property of a test for an infected individual and its revision should only happen when another infected individual panel with shorter period is available.

4.4.3 Delta-value

The delta-value example used a healthy subject sample of 257 blood donors' specimens with negative results for anti-HCV, ALT, and HCV RNA in the latest three blood donations and an infected individuals' samples of 17 infected individuals' specimens with confirmed HCV infection from a serum bank of the Portuguese Institute of Blood and Transplantation with positive tests results of immunoassay, recombinant immunoblot, and polymerase chain reaction, and twelve samples from commercial panels (6 samples from a qualification panel QHV711, and 6 samples from a genotype performance panel PHV350). The samples were measured by the same test, which is classified in this model as "test 1", and by another screening test for anti-HCV designated as "test 2". For the "test 1" $\delta+_{cand} = 3.831$ and for the "test 2" $\delta+_{comp} = 2.792$. From the considerations above, "test 1" has a lower probability of generating indeterminate results from the infected candidates' sample, i.e., higher chance of false negative results. For the "test 1" $\delta-_{cand} = 9.441$ and for the "test 2" $\delta-_{comp} = 9.404$. Despite "test 1" having a somewhat lower probability of generating indeterminate results from the healthy subject sample, i.e., higher chance of false positives results, these results do not show a statistically significant difference. Equating false positive results to budget waste, the tests are classified in the same rank. The measurement should be reviewed as stated for the precision model [25].

4.5. Conclusions

To evaluate the TAE and measurement uncertainty results, the laboratory staff must be understood the theoretical principles of both models. TAE could be an alternative to determine the "grey-zone" however the medical laboratory staff must recognize it as not an estimator of measurement uncertainty. So, the TAE results cannot be used as a substitution to the expanded uncertainty, since the concepts are associated to different metrology theories.

Since the binary result is influenced by the indeterminate interval, a tripartite reporting should be used to determine the window period. Therefore, it should be measured using also the indeterminate results when available, allowing the identification of a most accurate period without increasing risk to post-transfusion safety. Window period traduces a major component of biological bias and it is recommended to be reported together with measurement uncertainty or TAE.

The delta-value proved to be a useful model not only to measure the chance of false negative results but also to measure the chance of waste in blood establishment budget due to false positive results.

References

1. International Organization for Standardization. ISO 15189 Medical laboratories - Requirements for quality and competence. 2nd ed. Geneva: ISO, 2012.

2. Commission Directive 2002/98/EC of the European Parliament and of the Council as regards setting standards of quality and safety for the collection, testing, processing, storage and distribution of human blood and blood components and amending Directive 2001/83/EC. Official Journal of the European Union L33/30-40.

3. Commission Directive 2004/33/EC Implementing Directive 2002/98/EC of the European Parliament and of the Council as regards certain technical requirements for blood and blood components. Official Journal of the European Union L91/25-39.

4. Commission Directive 2005/61/EC Implementing Directive 2002/98/EC of the European Parliament and of the Council as regards traceability requirements and notification of serious adverse reactions and events. Official Journal of the European Union L256/32-40.

5. Commission Directive 2005/62/EC Implementing Directive 2002/98/EC of the European Parliament and of the Council as regards Community standards and specifications relating to a quality system for blood establishments. Official Journal of the European Union L256/41-48.

6. American Association of Blood Banks. Accreditation program. Retrieved from: http://www.aabb.org/sa/overview/Pages/program.aspx Accessed December 22, 2015.

7. European Blood Inspection System. Common European standards and criteria for the inspection of blood establishments. The European Union: EuBIS, 2010.

8. International Organization for Standardization. ISO 9001 Quality management systems - Requirements. 5th ed. Geneva: ISO, 2015.

9. Bureau International des Poids et Mesures. Evaluation of measurement data - Guide to the expression of uncertainty in measurement. JCGM 100:2008, GUM 1995 with minor corrections. Sèvres: BIPM, 2008.

10. Serology Uncertainty of Measurement Working Party. Uncertainty of measurement calculation for the Abbott Murex HCV EIA using the guide to the expression of uncertainty in measurement (GUM) approach. Version 2.0. Fitzroy, Vic: The Working Party; 2005.

11. Dimech W, Francis B, Kox J, Roberts G. Calculating uncertainty of measurement for serology assays by use of precision and bias (Technical brief). Clin Chem 2006;52(3):526-529.

12. Bureau International des Poids et Mesures. JCGM 200:2012 International vocabulary of metrology - Basic and general concepts and associated terms, 2008 version with minor corrections. 3rd ed. Sèvres: BIPM, 2012.

13. Westgard J. Managing quality vs. measuring uncertainty in the medical laboratory. Clin Chem Lab Med 2012, 48(1):31–40.

14. Pulido A, Ruisánchez I, Boqué R, Rius F. Estimating the uncertainty of binary tests results to assess their compliance with regulatory limits. Anal Chim Acta 2002, 455(2):267-275.

15. Pulido A, Ruisánchez I, Boqué R, Rius F. Uncertainty of results in routine qualitative analysis. Trends Anal Chem 2003, 22(10):647-654.

16. Ellison S. Uncertainties in qualitative testing and analysis. Accredit Qual Assur 2000, 5(8):346-348.

17. Mil'man B, Konopel'ko L. Uncertainty of qualitative chemical analysis: general methodology and binary test methods. J Anal Chem 2004, 59(12):1128-1141.

18. Van Hove L. A Six Sigma Operational Excellence Strategy Applied to Assay Design, Abbott Laboratories. Presented at the Six Sigma Quality Design and Control Workshop of Westgard QC, Jun 28-29, 2015; Madison (WI), USA.

19. Clinical and Laboratory Standards Institute. EP21-A Estimation of total analytical error for clinical laboratory methods, approved guideline. Philadelphia (PA): CLSI, 2003.

20. US Food and Drug Administration, Center for Devices and Radiological Health. Guidance for industry and FDA staff: recommendations for Clinical Laboratory Improvement Amendments of 1988 (CLIA) waver applications for manufacturers of in vitro diagnostic devices. Silver Spring (MD): FDA, 2008.

21. Westgard J, Neil Carey R, Wold S. Criteria for judging precision and accuracy in method development and evaluation. Clin Chem 1974, 20(7):825-833.

22. Eisenhart E. Realistic evaluation of the precision and accuracy of instrument calibration systems. J Res Nat Bur Stand 1963, Sect C 67C,161.

23. Mo M. Total analytical error evaluation in quantitative and qualitative assays. Presented at the FDA/Industry Statistics Workshop - Statistics in the FDA and Industry: Past, Present, and Future of Food and Drugs Administration Sep 27-29, 2006; Alexandria (VA), USA. Retrieved from http://www.amstat.org/meetings/fdaworkshop/presentations/2006/PS9_MayMo_Total%20Error_FDAInd%202006.ppt Accessed December 22, 2015.

24. European Co-operation for Accreditation. EA-4/02 M: 2013 Expression of the uncertainty of measurement in calibration. 2nd ed. Europe: EA, 2013.

25. Pereira P, Westgard J, Encarnação P, Seghatchian J. Analytical model for calculating indeterminate results interval of screening tests, the effect on seroconversion window period: a brief evaluation of theimpact of uncertain results on the blood establishment budget. Transfus Apher Sci 2014, 51(2):126-131.

26. International Organization for Standardization. ISO 5725-2 Accuracy (trueness and precision) of measurement methods and results - Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method. Geneva: ISO, 1994.

27. Le T, Bhushan V, Vasan N. First aid for the USMLE step 1. 20th ed. New York (NY): McGraw-Hill Medical, 2009.

28. SeraCare Life Sciences, Inc. Hepatitis C seroconversion panel (PHV901) HCV genotype 1a. Manufacturer data sheet. Milford, (MA): Seracare, 2006.

29. Clinical and Laboratory Standards Institute. M53-A Criteria for laboratory testing and diagnosis of human immunodeficiency virus infection; approved guideline. Philadelphia (PA): CLSI, 2011.

30. American Association of Blood Banks. Standards for blood banks and transfusion services. 29th ed. Bethesda (MD): AABB, 2014.

31. Fiebig E, Wright D, Rawal B, Garrett P, Schumacher R, Peddada L, et al. Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. AIDS 2003, 17(13):1871-1879.

32. Crofts N, Maskill W, Gust I. Evaluation of enzyme-linked immunosorbent assays: a method of data analysis. J Virol Methods 1988, 22(1):51-9.

33. Ellison SLR and Williams A. Use of uncertainty information in compliance assessment. Eurachem/CITAC, 2007.

34. Sandberg S. Summing up and further work. Presented at the 1st EFLM strategic conference: Defining analytical performance goals 15 years after the Stockholm conference, Nov 24-25, 2004; Milan, Italy. Retrieved from http://www.eflm.eu/files/efcc/21%20-%20Summary%20II%20Sandberg.pdf Accessed December 22, 2015.

Chapter 5 – Evaluation of the Diagnostic Uncertainty in Screening Immunoassays in Blood Establishments: Computation of Diagnostic Accuracy Models

This page intentionally left blank

5.1. Abstract

Measurement uncertainty is not intended to determine diagnostic accuracy of a test. Models focused on probability have been proposed in medical laboratories' screening tests. This Chapter reviews, discusses and proposes models to determine diagnostic accuracy and agreement of binary results in blood establishments' screening tests in a risk-based viewpoint. Example applications are provided for a test to measure the antibodies to the hepatitis C virus. The results show that these models satisfy ISO 15189 principles and the estimation of diagnostic sensitivity, diagnostic specificity, and area under the ROC curve allow also to determinate the uncertainty of these estimations which could be referred as "diagnostic uncertainty". The estimation of agreement of binary results allows to determine what could be referred as "agreement uncertainty".

5.2. Introduction

Tests with ordinal (entry 1.26 of [1]) results in medical laboratories are commonly referred to as "diagnostic tests" [2] or "qualitative tests" [3]. In blood establishments they are known as "screening tests" [4]. The definition of measurement uncertainty (entry 2.26 of [5]) limits its application to numerical quantity results; however medical laboratories already dealt with ordinal binary results (note: these results are also recognized in medical laboratory as "qualitative" or "semi-quantitative" results). See Chapter 1 for a discussion of ISO 15189 requirements to quality control and Chapter 2 for a discussion about other causes of uncertain results than measurement uncertainty.

In medical laboratory field, The Australian National Pathology Accreditation Advisory Council (NPAAC) guideline (2007) [6], and the Clinical and Laboratory Standards Institute (CLSI) EP29-A guideline (2012) recommend a set of GUM measurement uncertainty approaches [7], for what cannot be applied to the estimation of binary results' uncertainty.

CLSI defines "diagnostic accuracy" as the "the extent of agreement between the information from the test under evaluation and the diagnostic accuracy criteria" (entry 5.3 of [3]) or "the ability of a test system to obtain the correct result" (entry 1.4 of [8]). It can be estimated by diagnostic sensitivity and specificity pairs, and area under the receiver operating characteristic (ROC) curve. This Chapter discusses diagnostic

accuracy models, relating the results to the intended use of examination results, i.e., post-transfusion safety, and agreement of binary results when diagnostic is unknown.

The theoretical principles were implemented using standard spreadsheet software (Microsoft® Excel® 2016).

5.3. Methods and materials

5.3.1. Diagnostic accuracy I: 2x2 Contingency tables

Diagnostic accuracy methods measure the agreement between the screening test binary results and the diagnostic accuracy criteria (i.e., disease/non-disease). A Bayesian probability framework is adopted. The infected individuals' sample should be carefully selected to prevent statistically significant "spectrum bias", i.e., "bias between estimated test performance and true test performance when the sample used for evaluating an assay does not properly represent the entire disease spectrum over the target (intended-use) population" (entry 4.2 of [2]). When the diagnosis is unknown, a comparative method should be used to measure the degree of concordance between the screening test binary results and the binary results of a comparative test [3] (see 5.3.3).

The measurement of the percentage of true-positive results among the test results for a sample known to be positive for the test is known as "diagnostic sensitivity" (entry 5.3 of [3]) *se* [%], it is measured through the mathematical model $[TP / (TP + FN)] \cdot 100$, where *TP* is the number of true-positive results and *FN* is the number of false-negative results. The measurement of the percentage of false-negative results among the test results for a sample known to be positive for the test is known as "false negatives ratio" or could be referred as "diagnostic uncertainty", it is measured through the mathematical model $[FN / (TP + FN)] \cdot 100$. So, "diagnostic uncertainty" of a screening immunoassay, could be defined as "the risk of false results". Its concept is similar to measurement uncertainty, in this case measurement uncertainty of binary results. Since diagnostic sensitivity is equal to 1 - FN = $1 - \beta$ -error, it is analogous to statistical power of test (entry of 2.1 [9]). On the non-disease sample, the percentage of true-negative results among the test results for a sample known to be negative for the test is known as "diagnostic specificity" (entry 5.3 of [3]) *sp* [%], it is measured through the model $[TN / (FP + TN)] \cdot 100$, where *TN* is

the number of true-negative results and *FP* is the number of false-positive results. The percentage of false-positive results among the test results for a sample known to be negative for the test is known as "false positives ratio" or could be referred as "diagnostic uncertainty", it is measured through the mathematical model $[FP / (FP + TN)] \cdot 100$.

Transfusion safety requires that blood establishments' immunoassays should have a high diagnostic sensitivity to ensure that results from infected donors have a high probability to be expressed as "positive". Thus, the probability of generating false negative results (β-error or type II error) should be minimized, at the expense of having a larger probability of generating false positive results (α-error or type I error). The α and β errors are easily identified to the statistical hypothesis testing. Information about statistical hypothesis testing and diagnostic tests can be found elsewhere [9,10].

Table 1 2x2 Contingency table for diagnostic accuracy.

	Diagnostic accuracy criteria		
Candidate test results	Positive (Disease, $D = 1$)	Negative (Non-disease, $D = 0$)	Total
Positive ($y = 1$)	True-positive results (*TP*)	False-positive results (*FP*) α - error	$TP + FP$
Negative ($y = 0$)	False-negative results (*FN*) β - error	True-negative results (*TN*)	$FN + TN$
Total	$TP + FN$	$FP + TN$	N

Consider the case where samples of infected and uninfected blood are available from a set of infected individuals diagnosed with/without the disease or with positive/negative result from a reference or "gold-standard" test (note: there are no "gold-standard" tests for most viral agents tested in a blood establishments' laboratory). The infected sample should include specimens with all types and sub-types of agents with epidemiological prevalence in the geographic area of the blood donors; otherwise a bias effect is induced that tends to overestimate the test sensitivity. The uninfected sample should not include specimens from subjects with known interference factors (for example, pre-analytical effects such as anti-coagulant, bilirubin, erythrocytes, hemoglobin, dialysis, etc., disease effects, drugs' effects, herbs and natural products effects), as they might have an impact on the specificity estimate. The blood donors' population is a recommended source for this group since the donors have a long known medical screening record representing assurance of healthy condition. Test examinations should be performed under reproducibility conditions during 10-20 days.

Low and high limits for the 95 % confidence interval (95 %CI) of the test diagnostic sensitivity can be computed respectively from the formulas LL_{se} [%] = ($Q_{1,se}$ - $Q_{2,se}$) / $Q_{3,se}$ · 100 and HL_{se} [%] = ($Q_{1,se}$ + $Q_{2,se}$) / $Q_{3,se}$, where $Q_{1,se} = 2 \cdot TP + 1.96^2$, $Q_{2,se} = 1.96 \cdot \sqrt{1.96^2 + 4 \cdot TP \cdot FN / (TP + FN)}$, and $Q_{3,se} = 2 \cdot (TP + FN + 1.96^2)$. Low and high limits for the test diagnostic specificity are given respectively by *LLsp* [%] = ($Q_{1,sp}$ - $Q_{2,sp}$) / $Q_{3,sp}$ · 100, and HL_{sp} [%] = ($Q_{1,sp}$ + $Q_{2,sp}$) / $Q_{3,sp}$, where $Q_{1,sp} = 2 \cdot TN + 1.96^2$; $Q_{2,sp} = 1.96 \cdot \sqrt{1.96^2 + 4 \cdot FP \cdot TN / (FP + TN)}$, and $Q_{3,sp} = 2 \cdot (FP + TN + 1.96^2)$ [3].

Considering the confidence level equal to 1 – α-error, the significance level is equal to 5 %. In practice, it can be said that it is 95 % confident about the inclusion of the sample mean in the interval or the error margin of 5 % attributed to that statement. The significance level is analogous to “diagnostic uncertainty”.

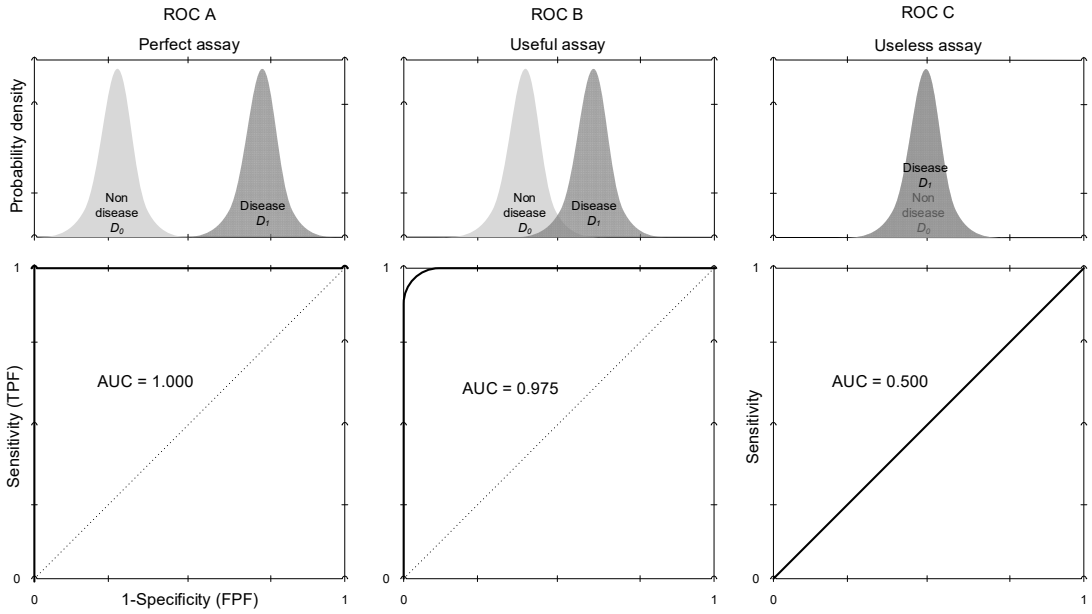
5.3.2. Diagnostic accuracy II: Receiver operating characteristic curve

CLSI defines a ROC curve as “a graphical description of test performance representing the relationship between the true-positive fraction (sensitivity) and the fraction (1 - specificity)” [2]. The ROC curve is simply a graphical plot displaying the performance of the diagnostic test as the “cutoff” varies. The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the test. The area under the curve (AUC) is a measure of the ability of the test to correctly classify positive and negative samples (accuracy measurement), often called discrimination. It ranges from 0.5, for a test that randomly assigns a positive or negative result to a sample, to 1, for a perfect test (sensitivity = 1 and 1 - specificity = 0). Figure 1 shows three ROC curves, where ROC A assures 100 % true results using infinite number of determinations from non-disease and disease samples which is unrealistic, ROC B assures a high probability of true results which is realistic, and ROC C assures 100 % true false results which is unrealistic. Parametric and non-parametric methods exist to compute AUC [9]. The D’Agostino-Pearson normality test [11] can be used to determine if test results distributions are indeed normal. When the test results of the positive samples and the negative samples have a normal distribution, AUC can be computed from $AUC_p = FPF(c) \cdot TPF(c)$, where $FPF(c)$ is the false positive fraction for a considered “cutoff” value c and $TPF(c)$ is the true positive fraction for the same “cutoff” value [12]. When the distributions of the test results are unknown, one can use the non-parametric *Mann-Whitney U* statistical test. Test results should be ranked and AUC is computed from $AUC_{np} = U / n_1 \cdot n_0$, where $U = (n_1 \cdot n_0 + n_0 \cdot (n_0 + 1)) / (2 - R)$ where R

is the rank sum of squares, n_1 the number of negative samples, and n_0 the number of positive samples [13]. The difference between one and the *AUC* could be referred as "diagnostic uncertainty".

Test discrimination is then classified from the AUC values according to the following scale [14]: if $AUC \in [0.50, 0.70[$ the discrimination is poor, if $AUC \in [0.70, 0.80[$ the discrimination is acceptable, if $AUC \in [0.80, 0.90[$ the discrimination is excellent, and if $AUC \in [0.90, 1.00]$ the discrimination is outstanding.

Figure 1 Three hypothetical ROC curves.



The measurement uncertainty of AUC is conditioned by the number and variance of true positive and false positive results measuring the dispersion of results in both samples. For example, a small number of true positive results or a large number of false positive results will increase the AUC measurement uncertainty. Also, lower numbers of true positive results due to sample size of infected individuals, will increase measurement uncertainty, i.e., decrease the confidence in the AUC outcome. DeLong *et al*. (1988) published a non-parametric model for the AUC standard error *SE* determination where $SE[AUC] = Var[AUC]$ which is computed by $Var[AUC] = Var[TP] / n_1 + Var[FP] / n_2$ where $Var[TP]$ is the variance of true positive results, n_1 is the number of true positive results, $Var[FP]$ is the variance of false positive results, n_2 is the number of false positive results [15]. The confidence interval must be associated with AUC to assess whether changes in AUC between classifiers are statistically significant. If only absolute

AUC value is known the evaluator could erroneously infer about the ability to separate the two samples. The 95 % CI for the AUC could be calculated as $AUC \pm 1.96 \cdot SE[AUC]$. Meanwhile, although $SE[AUC]$ has the apparent advantage of produce intervals centered on the absolute AUC result, Newcombe (1998) did not recommend its determination since the feature of this symmetry leads to defects of measurement, such as "overshoot" and "degeneracy". Newcombe recommended the Clopper-Pearson model for computing exact binomial confidence intervals which could be used for the 95 % CI determination eliminating both defects occurrence [16]. This model considers the interval [LL, HL] with $LL \leq p \leq HL$, such that for all θ in the interval: (a) if $LL \leq \theta \leq p$, $kp_r + \Sigma_{j:r<j\leq n} p_{j\geq\alpha/2}$; and (b) if $p \leq \theta \leq HL$, $kp_r + \Sigma_{j:0\leq j<n} p_j + kp_{r\geq\alpha/2}$where where $j = 0, 1, \ldots, n$, R denoting the random variable of which r is the realization, and $k = 1$ [17].

The significance level is equal to 5 % and it is analogous to "diagnostic uncertainty" (see 5.3.1).

5.3.3. Agreement of binary results when comparator is other than diagnostic accuracy criteria

When the diagnosis is unknown, the same ideas can be used to evaluate the test results agreement, however, the determinations are not intended to measure "diagnostic accuracy". Generally, this approach is suggested for "in house" and "non-waived" [18] modified tests when a true diagnosis is unavailable [3]. A comparative test for which diagnostic accuracy is known and with sensitivity and specificity values according to candidate test claim should be chosen, when available [19].

Table 2 summarizes the concepts of results agreement and the corresponding α and β errors when the diagnosis is unknown. The overall agreement OA [%] expresses the percentage of the positive and

Table 2 2x2 Contingency table for test results agreement.

	Comparative test		
Candidate test results	Positive ($x = 1$)	Negative ($x = 0$)	Total
Positive ($y = 1$)	a	b α - error	$a + b$
Negative ($y = 0$)	c β - error	d	$c + d$
Total	$a + c$	$b + d$	n

negative results agreement between the two tests and is given by $(a + d) / n \cdot 100$, where a is the number of candidate test positive results among the positive results in the comparative test, d is the number of

candidate test negative results among the negative results in the comparative test, and *n* is the number of samples. Low and high limits for overall agreement are computed, respectively, by LL_{OA} [%] = ($Q_{1,OA}$ - $Q_{2,OA}$) / $Q_{3,OA}$ · 100 and HL_{OA} [%] = ($Q_{1,OA}$ + $Q_{2,OA}$) / $Q_{3,OA}$ · 100, where $Q_{1,OA} = 2 \cdot (a + d) + 1.96^2$, $Q_{2,OA} = 1.96 \cdot \sqrt{1.96^2 + 4 \cdot (a + d) \cdot (b + c) / n}$, and $Q_{3,OA} = 2 \cdot (n + 1.96^2)$. The positive agreement *PA* [%] expresses the percentage of the positive results agreement between the two tests and is given by $a / (a + c) \cdot 100$. Low and high limits for positive agreement can be computed, respectively, from LL_{PA} [%] = ($Q_{1,PA}$ - $Q_{2,PA}$) / $Q_{3,PA}$ · 100 and HL_{PA}[%] = ($Q_{1,PA}$ + $Q_{2,PA}$)/$Q_{3,PA}$ · 100, where $Q_{1,PA} = 2 \cdot a + 1.96^2$, $Q_{2,PA} = 1.96 \cdot \sqrt{1.96^2 + 4 \cdot a \cdot c / (a + c)}$, and $Q_{3,PA} = 2 \cdot (a + c + 1.96^2)$. Finally, the negative agreement *NA* [%] expresses the percentage of the negative results agreement between the two tests and is given by $d / (b + d) \cdot 100$. Low and high limits for negative agreement are respectively given by LL_{NA} [%] = ($Q_{1,NA}$ - $Q_{2,NA}$) / $Q_{3,nA}$ · 100, and HL_{NA} [%] = ($Q_{1,NA}$ + $Q_{2,NA}$) / $Q_{3,NA}$ · 100, where $Q_{1,PA} = 2 \cdot d + 1.96^2$, $Q_{2,NA} = 1.96 \cdot \sqrt{1.96^2 + 4 \cdot b \cdot d / (b + d)}$, and $Q_{3,NA} = 2 \cdot (b + d + 1.96^2)$ (note: *a*, *b*, *c*, and *d* are respectively equivalent to diagnostic accuracy “true positive”, “false positives”, “false negatives”, and “true negatives”) [3].

The significance level is equal to 5 % and it is analogous to “diagnostic uncertainty” (see 5.3.1).

5.4. Results and discussion

To exemplify the models, a single screening anti-hepatitis C virus (HCV) immunoassay was used (Abbott Prism® HCV) (See Chapter 3 for details about the test) [20]. Indeterminate results cannot be considered in the estimations.

A diagnostic accuracy application for diagnostic sensitivity and diagnostic specificity determination was done using previous samples from infected and healthy individuals. Laboratory claimed the diagnostic sensitivity should be equal to 100 %, diagnostic specificity equal or higher than 98 %, and the lower limit for confidence intervals should be equal or higher than 85 % for sensitivity and equal or higher than 90 % for specificity. The estimations of test sensitivity and specificity are shown in Table 3, illustrating that the performance requirements have been achieved. According to Table 3, both were equal to 0 % to diagnostic sensitivity and clinical specificity and can be associated uniquely with the tested samples.

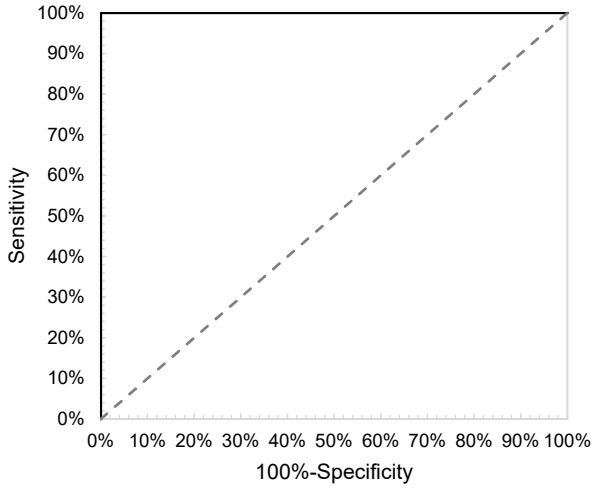
Table 3 Sensitivity and specificity results for an anti-HCV antibodies screening test.

	95 % CI
Se = 29 / (29 + 0) · 100 = 100 %	LL_{Se} = (61.84 - 3.84) / 65.68 = 88.3 %
	HL_{Se} = (61.84 + 3.84) / 65.68 · 100 = 100 %
Sp = 253 / (0 + 253) · 100 = 100 %	LL_{Sp} = (509.84 - 3.84) / 513.68 · 100 = 98.5 %
	HL_{Sp} = (509.84 + 3.84) / 513.68 · 100 = 100 %

Diagnostic sensitivity and specificity computed from Bayesian probabilities are the most accepted statistical tools applied to method validation for blood establishments' screening immunoassays since they have a critical role, providing 95 % CIs for the probabilities of occurrence of true results. These confidence intervals can be interpreted as the diagnostic accuracy interval. The drawback of the Bayesian probabilities approach is its susceptibility to inappropriate samples. For example, a small number of samples lead to confidence intervals with a relatively big amplitude due to the decrease of the statistical power, incorrectly classified samples in the supposedly group of healthy samples induce errors in the specificity estimate, an infected sample that is not representative of the agent prevalence in the population will induce a spectrum bias in the estimates. Since the prediction value of a negative, positive or both in a sample or population has interest in the clinical field, physicians may wish to know the probabilities of true and/or false results, and these can be easily provided for research. The determination should occur when significant epidemiological changes happen. In blood establishments the claimed diagnostic accuracy should be according to the published screening test reference values of sensitivity and specificity. When a medical laboratory claims 95 % CI minimum requirement, it must consider the available size of samples to determine diagnostic accuracy, since it is a limitation for the claim. For example, if the number of infected individuals' sample is 29, the best 95 % CI is 88.3-100 %; therefore, claims higher than 88.3 % are unrealistic. For the results agreement model, the same samples were used as for the diagnostic accuracy model where the comparative test gave 30 positive and 285 negative results, i.e., the sample is reclassified according to comparative test output.

For the ROC/AUC example, the same data were used as for the diagnostic accuracy model. Figure 2 shows the ROC curve for the HCV screening test using the same samples considered in the example of diagnostic accuracy. The curve suggests that the test is able to accurately predict the value of the sample. The perfect

Figure 2 ROC curve for an anti-HCV antibodies screening test with an AUC = 1.00.



test point (0, 1.00) is achieved with "cutoff" = 1.00. The AUC was computed using the nonparametric method described above, since the hypothesis of the normal distribution of test results was rejected by the D'Agostino-Pearson test. A value of 1.00 was obtained and thus the test ranks as an "outstanding

Table 4 Diagnostic sensitivity and specificity results of an anti-HCV antibodies screening test for 11 clinical decision points.

				Specificity			Sensitivity			
					95 % CI			95 % CI		
Percentil	>Cutoff	ND[a]	D[b]	Absolute (%)	LL[c] (%)	HL[d] (%)	Absolute (%)	LL (%)	HL (%)	1-Sp (%)
0	0.03	0	281	0.0	0.0	1.5	100.0	88.3	100.0	100.0
10	0.06	9	251	3.9	2.1	7.2	100.0	88.3	100.0	96.1
20	0.08	56	170	28.4	22.6	35.1	100.0	88.3	100.0	71.6
30	0.08	56	170	28.4	22.6	35.1	100.0	88.3	100.0	71.6
40	0.09	112	120	55.2	48.3	61.9	100.0	88.3	100.0	44.8
50	0.09	112	120	55.2	48.3	61.9	100.0	88.3	100.0	44.8
60	0.1	162	79	76.4	70.3	81.6	100.0	88.3	100.0	23.6
70	0.1	162	79	76.4	70.3	81.6	100.0	88.3	100.0	23.6
80	0.11	203	55	88.6	83.9	92.1	100.0	88.3	100.0	11.4
90	1.31	253	29	100.0	98.5	100.0	100.0	88.3	100.0	0.0
100	8.82	281	0	100.0	98.5	100.0	0.0	0.0	12.1	0.0

[a]Non-disease sample; [b]Disease sample; [c]Low-limit; [d]High-limit

discrimination" test. Following Newcombe recommendation, only the Clopper-Pearson model was considered as the recommended tool for measurement uncertainty on the AUC result. Performance considered was [0.90, 1.00]. The AUC 95 % CI was [0.99, 1.00]. The confidence interval was accepted since it is contained by the claimed interval. The low limit is close to 1.00, which strengthens the power of the

AUC absolute value. Table 4 shows the estimated sensitivity and specificity for a range of fixed and pre-specified “cutoff” points equivalent to the 0-100th percentile in steps of 10th percentile with the corresponding “criterion values”. The 95 % CI values have been calculated using the same mathematical models as for the sensitivity and specificity estimations. These calculations are also useful for the definition of “cutoff” point in a test under development. For example, a new blood establishment’s screening test should not have a significant lack of diagnostic sensitivity since it increases the risk of post-transfusion infection. The measurements necessary to plot the ROC curve are essential for IVD medical devices manufacturers to determine the “cutoff” value for a test under development, such as for medical laboratory “in house” and “non-waived” modified tests where the “cutoff” must be identified. The AUC single value can be applicable only to the infected and healthy individuals’ samples and it cannot be inferred to the population groups. For example, the “cutoff” point with a result of *AUC* = 1.00 separates perfectly the two groups, but it is unknown if it separates the two groups in population. Then, the AUC 95 % CI should be determined since it is an inference to population under same characteristics than the tested samples. When the population has different characteristics than the samples, the 95 % CI will be biased. Blood establishments’ virology laboratories using commercial tests cannot change cutoff. In this condition, the ROC curve utility is limited to display the test diagnostic accuracy, and test AUC depending on the sample used.

Considering the overall agreement, positive agreement, and negative agreement shall be equal or higher than 98 %, the overall agreement and negative agreement confidence interval low limit shall be equal or higher than 95 % and the positive agreement confidence interval low limit shall be equal or higher than 85 %. Table 5 shows the estimations of results agreement and indicates that the performance requirements have been met. The results agreement of a candidate and a comparative tests pair should be only determined when diagnosis is unknown. The considerations regarding agreements must to be taken with care, since there they can be misunderstood. The 95 % CI is the agreement accuracy. Such as in the diagnostic accuracy estimations, the confidence intervals can be interpreted as the agreement accuracy interval, and the significance level the 5 % risk of incorrectly rejecting the null hypothesis and can be interpreted as the “agreement uncertainty”. The performance requirements should follow the

Table 5 Agreement results for an anti-HCV antibodies screening test.

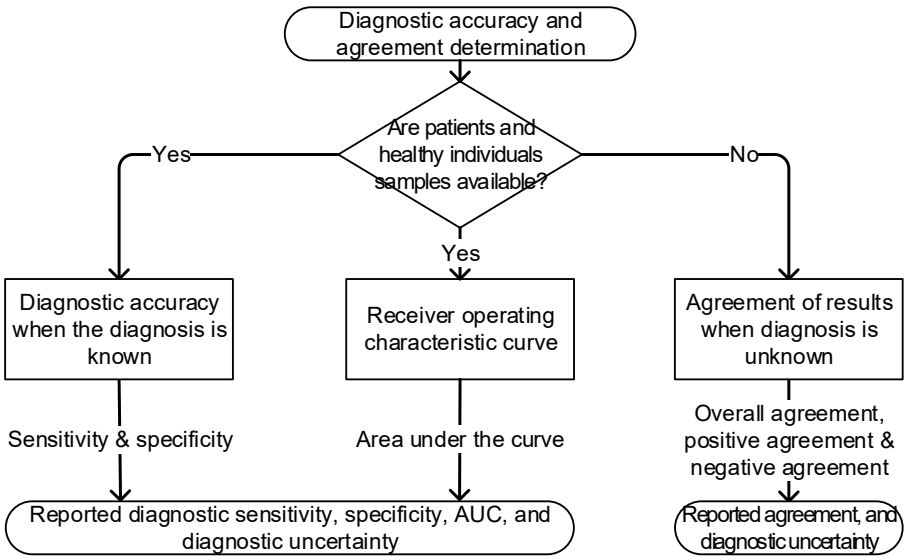
	95 % CI
OA = (28 + 253) / 282 · 100 = 99.6 %	LL_{OA} = (565.8 - 5.5) / 571.7 · 100 = 98.0 %
	HL_{OA} = (565.8 + 5.5) / 571.7 · 100 = 99.9 %
PA = 28 / (28 + 0) · 100 = 100 %	LL_{PA} = (59.84 - 3.84) / 63.68 · 100 = 87.9 %
	HL_{PA} = (59.84 + 3.84) / 63.68 · 100 = 100 %
PN = 253 / (253 + 1) · 100 = 99.6 %	LL_{NA} = (509.84 - 5.48) / 515.68 · 100 = 97.8 %
	HL_{NA} = (509.84 + 5.48) / 515.68 · 100 = 99.9 %

same principles than the diagnostic accuracy model. This model is not recommended to be used for commercial screening immunoassays.

A lack of screening immunoassays is the absence of metrological traceability/traceability chain due to the unavailability of reference methods or certified reference materials (entry 2.41 of [1]). Consequently, blood establishments' screening immunoassays are classified as untraceable tests. For a thorough discussion of traceability in medical laboratories, further information can be found elsewhere [21].

Figure 3 summarizes the models in a decision flowchart to choose between diagnostic accuracy models and agreement model.

Figure 3 Scheme for the selection of diagnostic accuracy, agreement and area under the receiver operating characteristic curve models for a blood establishments' screening test.



5.5. Conclusions

In summary, the diagnostic accuracy of blood establishments' screening immunoassays characterizes the level of accuracy of binary results, and false results ratio and the significance level of a confidence interval reflect the diagnostic uncertainty. Diagnostic accuracy allows the blood establishments to verify the claim of performance requirements focusing on the risk of false results and its effect in post-transfusion safety.

References

1. Bureau International des Poids et Mesures. JCGM 200:2012 International vocabulary of metrology - Basic and general concepts and associated terms, 2008 version with minor corrections. 3rd ed. Sèvres: BIPM, 2012.

2. Clinical and Laboratory Standards Institute. EP24-A2 Assessment of the clinical accuracy of laboratory tests using receiver operating characteristic curves. 2nd ed. Wayne (PA): CLSI, 2011.

3. Clinical and Laboratory Standards Institute. EP12-A2 User protocol for evaluation of qualitative test performance. 2nd ed. Wayne (PA): CLSI, 2008.

4. Clinical and Laboratory Standards Institute. MM10-A Genotyping for infectious diseases: identification and characterization. Wayne (PA): CLSI, 2006.

5. Bureau International des Poids et Mesures. Evaluation of measurement data - Guide to the expression of uncertainty in measurement. JCGM 100:2008, GUM 1995 with minor corrections. Sèvres: BIPM, 2008.

6. National Pathology Accreditation Advisory Council. Requirements for the estimation of measurement uncertainty. Barton, ACT: NPAAC, 2007.

7. Clinical and Laboratory Standards Institute. EP29-A Expression of measurement uncertainty in laboratory medicine. Wayne (PA): CLSI, 2012.

8. Clinical and Laboratory Standards Institute. MM03 Molecular diagnostic methods for infectious diseases. 3rd ed. Wayne, (PA): CLSI, 2015.

9. Zhou X, Obuchowski N, McCLish D. Statistical methods in diagnostic medicine, 2nd ed. Hoboken (NJ): John Wiley & Sons, 2002.

10. Pepe M. The Statistical evaluation of medical tests for classification and prediction. Oxford: Oxford University Press, 2004.

11. D'Agostino RB, Belanger A, D'Agostino RB Jr. A suggestion for using powerful and informative tests of normality. Am Stat 1990, 44(4):316-321.

12. Metz C. Basic principles of ROC analysis. Semin Nucl Med 1978, 8(4):283-298.

13. Mann H, Whitney D. On a test of whether one of two random variables is stochastically larger than the other. Ann Math Stat 1947, 18(1):50-60.

14. Hosmer D Jr, Lemeshow S, Sturdivant R. Assessing the fit of the model. In: Hosmer D Jr, Lemeshow S, Sturdivant R, editors. Solutions manual to accompany applied logistic regression. 2nd ed. Hoboken (NJ): John Wiley & Sons, 2013; p. 153-226.

15. DeLong E, DeLong D, Clarke-Pearson D. Comparing the area under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988, 44(3):837-845.

16. Newcombe R. Two sided confidence intervals for the single proportion: a comparative evaluation of seven methods. Stat Med 1998, 17(8):857-872.

17. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 1934, 26(4):404-413.

18. US Government. Laboratory requirements, 42 CFR Part 493 Subparts H, J, K and M, 2003.

19. US Food and Drug Administration, Center for Devices and Radiological Health. Guidance for industry and FDA staff: recommendations for Clinical Laboratory Improvement Amendments of 1988 (CLIA) waver applications for manufacturers of in vitro diagnostic devices. Silver Spring (MD): FDA, 2008.

20. Abbott Diagnostics Division. Abbott Prism HCV. Code: 84-5342/R9. Wiesbaden: Abbott Laboratories, 2008.

21. Vesper H, Thienpont L. Traceability in laboratory medicine. Clin Chem 2009, 55(6):1067-1075.

Part III - Conclusions

This page intentionally left blank

Chapter 6 - Scheme for the Selection of Models to the Estimation of Uncertainty in Blood Establishments' Screening Immunoassays

This page intentionally left blank

6.1. Abstract

Blood establishments routinely perform screening immunoassays to assess safety of the blood components. As with any other screening test, results have an inherent uncertainty. In blood establishments the major concern is the chance of false negatives, due to its possible impact on health of patients (blood receptor). This Chapters summarizes in a flowchart a set of models recommended to the estimation of “grey-zone”, impact of the number of indeterminate results in budget, seronegative period, diagnostic accuracy and agreement of binary results. The scheme is intended to support the screening laboratories’ staffs on the selection of a model to the uncertainty estimation according to the intended use of the tests’ results (post-transfusion safety). The discussion is risk-based, from the blood donor selection to the report of screening immunoassays’ results.

6.2. Introduction

The effect of the risk caused by the epidemiological prevalence of transmissible diseases on post-transfusion infection is already considered in blood establishments on the screening’s interview in candidates to blood donors. The risk of uncertain/false results on screening tests is also of major importance for blood establishments, as evidenced, for example, by cases where nucleic acid testing gives a false negative output due to virus mutations [1,2]. Risk assessment in blood establishments has been gaining importance on the fulfillment of ISO quality management systems [3], and on the satisfaction of the European Blood Inspection System and the European Directorate for the Quality of Medicines & HealthCare (EDQM) claims [4-6]. Currently, risk evaluation is not required by the European Union regulatory Directives [7-10] or by the American Association of Blood Banks Standards for Blood Banks [11]. However, a large number of European blood establishments are ISO 9001 certified, and the current edition requires “risk-based thinking” in decisions [4]. Measurement uncertainty (entry 2.26 of [12]) evaluation using an appropriate model is a step toward risk [entry 3.7.9 of (4)] evaluation.

A shorter number of blood establishment has screening immunoassays accredited according to ISO 15189. This international standard requires measurement uncertainty determination in all the laboratory tests (despite the output to be numerical, ordinal or nominal) [13].

The Guide to the Expression of Uncertainty in Measurement (GUM) [14] is considered by the metrology global organizations as the seminal guide to evaluate measurement uncertainty. However, the application of its models to the screening tests is not a current practice in blood establishments, not only because it is not systematically applied in medical laboratory, but also to be intended to numerical results. Alternatives to GUM models for uncertainty determination, focusing on diagnostic accuracy, are usually preferred [15,16]. However, the clinical decision point in screening immunoassays' ordinal scale is a numerical quantity, for what the measurement uncertainty close to this value ("grey-zone") should be estimated since it determines the binary results' accuracy.

6.3. Methods and materials

6.3.1. Measurement uncertainty and total analytical error

The International Vocabulary of Metrology (VIM) considers the uncertainty arising from the variance of measurement. GUM proposed a methodology that is widely used in chemistry and physics, but that is rarely considered in blood establishments' screening laboratories, such as in other medical laboratories. GUM applicability to screening tests' results producing an ordinal binary result (positive/negative) is questionable, but GUM models could provide to be useful to estimate measurement uncertainty at the "cutoff" using the principles currently recognized in metrology (Uncertainty Approach). Results equal or close to the "cutoff" area rapidly identified by blood establishment's staff as results with statistical significant risk to be false.

GUM models could be divided into modeling and empirical. The modeling models are mainly based on a model equation requiring a specific mathematical model (based in for each screening test respecting the stoichiometry of the test's reaction. This model combines the major measurement uncertainty components into a combined standard uncertainty using the law of the propagation of uncertainty, partial derivative method [14]) or the propagation of distributions, Monte Carlo simulation method [17]). The propagation of variance rules used in the model assume that all the sources of uncertainty are uncorrelated, which may not be true, thus leading to an over-estimation of the measurement uncertainty. The modeling estimations require mathematical and statistical skills which are usually unavailable in blood

establishments. Anyway, any model equation to describe the measurement is unknown, and therefore the modelling approach cannot be applied.

The empirical models are mainly based on experimental data coming from single-laboratory validation (including quality control), interlaboratory comparisons, and external quality assessment (EQA) [proficiency testing (PT)]. Any of three models can be used. In the single laboratory (intralaboratory) and in laboratory intercomparisons (interlaboratory) model model, within-laboratory reproducibility standard deviation s_{Rw} and bias uncertainty u_b are the components of combined standard uncertainty u_c. The within-laboratory reproducibility standard deviation could be determined from method validation or from internal quality control data, respectively. In the EQA/PT model the combined standard uncertainty is equal to group's repeatability standard deviation s_r [18]. Measurement uncertainty should be estimated with results close or equal to the "cutoff" (ratio equal to one). Estimations not close to 1.00 will have lower statistical significance/significant risk to be unrealistic. A nonhomogeneous EQA/PT group of laboratories may lead to an unrealistic estimation, for what this model should be considered uniquely if the two others are not feasible. Pereira et al. (2015) recommend the use of the intralaboratory approach using data of the Clinical Laboratory Standards Institute (CSLI) protocol EP15 as an initial uncertainty estimation in a new test. Since long-term data are available, it is recommended to apply the use of the intralaboratory approach using repeatability from ratio 0.5 to 1.5 and between-run "cuttoff" precision [19]. The output of the GUM models is an uncertainty interval, usually referred to as the "expanded uncertainty", generally representing a 95 % confidence interval (CI) for the average of the test results for a "cutoff" sample.

The blood establishments must then define the "rejection zone" [20]. Pereira et al. (2015) demonstrated the estimation of a decision interval by subtracting $1.65u_c$ from the "cutoff" ratio value [19]. The use of a "rejection zone" ("grey-zone") implies the use of ternary results, i.e., positive/indeterminate/negative. In blood establishments, the "rejection zone" has the practical effect of reducing the chance of binary results to be classified as false. The blood components with indeterminate results shall be eliminated. For a detailed discussion of GUM models in blood establishments' screening tests please refer to Chapter 3.

Total analytical error (TAE) show to be an alternative model to determine the “grey-zone”, however it cannot fulfill measurement uncertainty requirement since it is not according to the Uncertainty Approach principles [14]. Pereira et al. (2014) demonstrated in a case that the TAE calculation is impractical when bias cannot be determined, being equal to the within-laboratory reproducibility [21]. For further details, please refer to Chapter 4.

6.3.2. Seroconversion window period

The window period of an infectious disease test is the time between the first day of infection and the day when the test can reliably detect the infection [22]. The seroconversion window period expresses the major component of diagnostic bias and is part of the residual risk (entry 2.29 of [23]) of post-transfusion infection. Samples from infected individuals on window period are seronegative, i.e., false negative results. Consequently, the determination of the window period is complementary to the measurement uncertainty determination. Pereira et al. (2014) presented a novel recommendation to window period determination considering the use of a “rejection zone” to classify the test results, whereby a new definition is proposed: “the window period for a test designed to detect a specific disease (particularly an infectious disease) is the time between the first day of infection and the day when the test result cannot reliably rule out the possibility of infection (due to indeterminate results)” [21]. This change on the classical concept has no impact on post-transfusion residual risk, since blood donations with indeterminate results are excluded. Note that the window period gives a measure of the diagnostic bias, which is cannot be addressed by measurement uncertainty. For further details, please refer to Chapter 4.

6.3.3. The delta value

The delta value published by Crofts (1988) is primarily intended for the evaluation of the chance of immunoassays outcome to be close to the “cutoff” point [24]. The delta results are measured for the infected individuals’ sample, yielding a delta plus value, and for the healthy individuals’ sample, yielding a delta minus value. They are an estimate of the chance of getting indeterminate results in a series of immunoassays. Pereira et al. (2014) suggested the use of delta values as a complementary tool to measurement uncertainty, since they provide information on which tests from a series have the lower or the higher chance to constitute a waste of the blood establishment budget due to the cost of

complemental tests used to know the true result, the inherent cost of eliminated blood components, or the suspension of further blood donations, depending on the algorithm in use [21]. For further details, please refer to Chapter 4.

6.3.4. Diagnostic accuracy models

Diagnostic accuracy models aim at determining the chance of obtaining a true result. Though this probability is affected by analytical components, it is mainly determined by biological factors, such as the window period, and by the number of samples tested. Diagnostic accuracy models use 2 x 2 contingency tables and Bayesian probabilities to estimate the diagnostic sensitivity and specificity from tests results of infected and uninfected samples, respectively.

Another diagnostic accuracy model is the receiver operating characteristic (ROC) curve which graphically displays the performance of a screening test. It represents the relationship between the sensitivity and the false-positive ratio (equal to one minus the sensitivity) for a specific "cutoff" level. The area under the ROC curve (AUC) measures the ability of a screening test to correctly classify true results (accurate results) ranging from 0.5 (50/50 chance to be true) to 1 (true value) [16].

Diagnostic accuracy methods express the probability of obtaining a true (accurate) result by an absolute value or by a 95 % confidence interval (95 % CI) [15,16]. The absolute values are exclusively related to the tested samples and cannot be inferred to the populations of infected or healthy individuals. The 95 % CI represents the inference to the population characterized by sample attributes (for example, subtypes of a virus). This is a critical advantage when compared to the use of the absolute value.

The estimation of false results' ratio (positive or negative) could be referred as "diagnostic uncertainty". The 95 % CI can be interpreted as the diagnostic accuracy interval. Considering the confidence level equal to $1 - \alpha$-error, the significance level is equal to 5 %. The significance level is analogous to diagnostic uncertainty. Likewise, the difference between one and the AUC could be referred as diagnostic uncertainty, and the significance level equal to 5 % is analogous to diagnostic uncertainty. Pereira et al. (2015) had already considered the uncertainty concept applied to diagnostic accuracy models [25]. For further details, please refer to Chapter 5.

6.3.4. Agreement of binary results

When the diagnosis is unknown the agreement between results of the test under evaluation and a second test should be determined, being the secondary test output taken as the reference value (entry 5.18 of [12]). Both samples with positive and negative results are used to measure the agreement of results (overall agreement), the agreement of positive results and the agreement of negative results [15]. It is not possible in this case to estimate the diagnostic accuracy. The diagnostic uncertainty could be interpreted likewise to the diagnostic accuracy models. Pereira et al. (2015) had already recognized the uncertainty concept applied to agreement of binary results models [25]. For further details, please refer to Chapter 5.

6.4. Results and discussion

Table 1 shows the 95 % CIs obtained using GUM and TAE models on an immunoassay for the measurement of the concentration of antibodies to the hepatitis C virus (HCV) (Abbott Diagnostics, Abbott Park, IL, USA). Details on these results can be found in Chapter 3. In the case of screening immunoassay any model equation to describe the measurement is unknown and therefore the modeling approach was impractical. GUM modeling approach is well suited for reagent research and development (R&D), at the manufacturer level, since they rely on the estimation and evaluation of the most critical uncertainty components.

GUM empirical intralaboratory application in blood establishments is straightforward with realistic intervals, requiring uniquely data that are already available taken from method validation, internal quality control and External quality assessment (EQA) [proficiency testing (PT)]. The obtained confidence intervals should inform the evaluation of the “rejection zone” on the “cutoff” point (which should contain the confidence interval). Numerical results inside this zone must be interpreted as “indeterminate results”, requiring further tests to determine the true result. The EQA/PT (interlaboratory approach) measurement uncertainty reliability is lower than intralaboratory estimations since the degrees of freedom are lower, and also due to the lack of repeatability on results close to cutoff. Consequently, the intralaboratory approaches should be preferred, starting with the “Validation (CLSI EP15) data” approach to a new test, and as soon as reliable quality control data is available, to determine using the “Validation and QC data”

Table 1 Expanded uncertainty intervals close to the "cutoff" point (95 % CI) determined on an anti-HCV immunoassay (Abbott Diagnostics, Abbott Park, IL, USA).

Uncertainty Approach (U)			Error Approach (TAE)
Intralaboratory		EQA	Total analytical error
Validation (CLSI EP15) data	Validation and QC data	Group data	Validation and QC data
$u_c(y) = \sqrt{s_{Rw}^2 + u_b^2}$	$u_c(y) = \sqrt{s_{Rw}^2 + u_b^2}$	$u_c(y) = \sqrt{s_{Rw}^2}$	$TAE = b + k \cdot s_{Rw}$
21 %	36 %	28 %	39 %

approach. Comparing the TAE result, 39 %, with GUM intralaboratory "Validation and QC data" approach expanded uncertainty, 36 %, using the same within-laboratory reproducibility s_{RW}, and bias *b* data; TAE is overestimated and otherwise will be underestimated. For the measurement uncertainty estimation, the TAE cannot be used since its concept is not equivalent to measurement uncertainty. A TAE determination and expanded uncertainty are not equal estimations, differing on the principle of each one. Thereby, when required, measurement uncertainty must be determined using models fulfilling Uncertainty Approach theory associated to GUM. The decision limit of the "rejection zone" is calculated by subtracting 1.65·*u* from the "cutoff" ratio value of 1.00. According to Table 1, the decision limit is 0.69 [20] (see Chapter 3 and Chapter 4 for details of the calculations).

The window period using the proposed definition was equal to 97 days. In the panel tested there was no indeterminate results, for what the 97th day was the first with a positive result. The use of the indeterminate result in window period will give a more realistic estimation. Details on the computation of the results can be found in Chapter 4.

The delta value results showed a close chance of the "test 1" and a second test to produce indeterminate results in healthy individuals' samples (9.441 ≅ 9.404) and a small higher chance of the "test 1" to produce indeterminate results in infected individuals' samples (3.831 > 2.792), translating also a smaller chance of waste in the budget. It could be a useful complementary tool to measurement uncertainty to a brief evaluation of the impact of indeterminate results in blood establishments' budget. See Chapter 4 for details of the calculations.

The diagnostic accuracy tests unlike measurement uncertainty, is well recognized in blood establishments. Table 2 shows results obtained for diagnostic accuracy and agreement of binary results. Even with 100 %

of true positives and true negatives, the diagnostic sensitivity and specificity interval is limited, respectively, by the number of infected individuals and the number of healthy individuals. This must be taken into consideration when evaluating these confidence intervals, since it does not mean a lack of accuracy but a significant lower confidence of the intervals' interpretation. The effect of the samples' sizes is not critical in the same way, since the AUC method is oriented to the ability of the test to classify true results on a "cutoff" point. An $AUC \in [0.99, 1.00]$ suggests an outstanding discrimination of the results, i.e., high chance of accurate classification, low chance of inaccurate classification. Details on the computation of the results can be found in Chapter 5.

Table 2 Number of days of the seroconversion window period, diagnostic accuracy and agreement of binary results

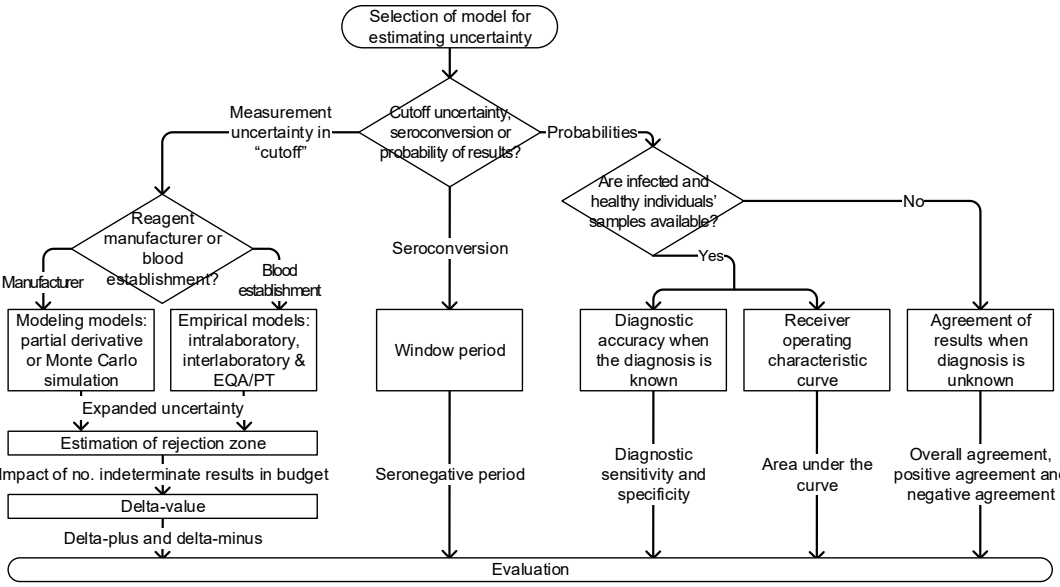
		Diagnostic accuracy			Agreement		
	Seroconversion window period (days)	Diagnostic sensitivity (%)	Diagnostic specificity (%)	Area under the curve (%)	Overall (%)	Positive (%)	Negative (%)
Absolute result	97[a]	100	100	100	99.6	100	99.6
95 % CI	NA	[88.3, 100]	[98.5, 100]	[99, 100]	[98, 99.9]	[87.9, 100]	[97.8, 99.9]

[a] PHV901 panel (SeraCare Life Sciences, Inc., Massachusetts); NA, not applicable

The agreement of result intervals is also limited by the sample sizes and the by number of true results. In the example in Table 2, the negative and overall agreements were limited mainly by the one non-compliant result in the healthy individuals' sample. The intervals under the agreement methods were computed taking as reference values the values given by a comparative test when there is no guarantee that they are true results. This model is thus not recommended to be used for commercial screening immunoassays in blood establishments. It might be useful only when infected and healthy individuals' samples are unavailable, something that is not expected in blood establishments.

To satisfy guidelines' claims, blood establishments must choose a model or models to uncertainty estimation and evaluation. Following the discussion above, several methods can be applied depending on the goal (for example, estimation of a "rejection zone" or diagnostic accuracy), on where the method is being used (manufacturer or blood establishment level), and the data available. The model selection can be made using Figure 1, according to the novel flowchart proposed by Pereira et al. (2015) [26].

Figure 1 Flowchart for the selection of uncertainty models in blood establishments' screening laboratories.



6.5. Conclusions

The uncertainty in screening immunoassays in blood establishment must be understood according to different concepts of uncertainty. Measurement uncertainty is one, but others are critical to the post-transfusion safety, such as the diagnostic accuracy models. GUM models applied to results near the "cutoff" concentration were demonstrated to be suitable methods to evaluate measurement uncertainty and diagnostic accuracy models demonstrated to be complementary. While measurement uncertainty provides the evaluation of the test "rejection zone"; diagnostic sensitivity, specificity and AUC permit for the classification of a test according to its probability of yielding true results. Exceptionally, when diagnosis is unknown, the agreement of binary results estimation is suitable. Window period is part of the residual risk and its determination is complementary not only to measurement uncertainty but also to diagnostic accuracy. Delta-value could be useful for a simple estimation of the impact of indeterminate results' number in budget. Based on the features of each available method to evaluate uncertainty, the better suited method(s) for risk assessment of screening immunoassays' results in blood establishments can be recommended.

This page intentionally left blank

References

1. Müller B, Nübling C, Kress J, Roth W, De Zolt S, Pichl L. How safe is safe: New human immunodeficiency virus type 1 variants missed by nucleic acid testing. Transfusion 2013, 53(10):2422-2430.

2. Pereira P, Westgard J, Encarnação P, Seghatchian J, de Sousa G. The role of uncertainty in results of screening immunoassays in blood establishments. Transfus Apher Sci 2015, 52(2):252-255.

3. International Organization for Standardization. ISO 9001 Quality management systems - Requirements. 5th ed. Geneva: ISO, 2015.

4. Pereira P, Westgard J, Encarnação P, Seghatchian J, Sousa G. Quality management in the European screening laboratories in blood establishments: a view on current approaches and trends. Transfus Apher Sci 2015, 52(2):245-251.

5. European Blood Inspection System. Common European standards and criteria for the inspection of blood establishments. The European Union: EuBIS: 2010.

6. European Directorate for the Quality of Medicines & HealthCare. Guide to the preparation, use and quality assurance of blood components. 18th ed. The European Union: EDQM, 2015.

7. Commission Directive 2002/98/EC of the European Parliament and of the Council as regards setting standards of quality and safety for the collection, testing, processing, storage and distribution of human blood and blood components and amending Directive 2001/83/EC. Official Journal of the European Union L33/30.

8. Commission Directive 2004/33/EC Implementing Directive 2002/98/EC of the European Parliament and of the Council as regards certain technical requirements for blood and blood components. Official Journal of the European Union L91/25-39.

9. Commission Directive 2005/61/EC Implementing Directive 2002/98/EC of the European Parliament and of the Council as regards traceability requirements and notification of serious adverse reactions and events. Official Journal of the European Union L256/32-40.

10. Commission Directive 2005/62/EC Implementing Directive 2002/98/EC of the European Parliament and of the Council as regards Community standards and specifications relating to a quality system for blood establishments. Official Journal of the European Union L256/41-48.

11. American Association of Blood Banks. Standards for blood banks and transfusion services. 29th ed. Bethesda (MD): AABB, 2014.

12. Bureau International des Poids et Mesures. JCGM 200:2012 International vocabulary of metrology - Basic and general concepts and associated terms, 2008 version with minor corrections. 3rd ed. Sèvres: BIPM, 2012.

13. International Organization for Standardization. ISO 15189 Medical laboratories - Requirements for quality and competence. 2nd ed. Geneva: ISO, 2012.

14. Bureau International des Poids et Mesures. Evaluation of measurement data - Guide to the expression of uncertainty in measurement. JCGM 100:2008, GUM 1995 with minor corrections. Sèvres: BIPM, 2008.

15. Clinical and Laboratory Standards Institute. EP12-A2 User protocol for evaluation of qualitative test performance. 2nd ed. Wayne (PA): CLSI, 2008.

16. Clinical and Laboratory Standards Institute. EP24-A2 Assessment of the clinical accuracy of laboratory tests using receiver operating characteristic curves. 2nd ed. Wayne (PA): CLSI, 2011.

17. Bureau International des Poids et Mesures. JCGM 101 Evaluation of measurement data - Supplement 1 to the “Guide to the expression of uncertainty in measurement” - Propagation of distributions using a Monte Carlo method. Sèvres: BIPM, 2008.

18. European Federation of National Associations of Measurement, Testing and Analytical Laboratories. Technical Report No. 1/2007 Measurement uncertainty revisited: Alternative approaches to uncertainty evaluation. EuroLab, 2007.

19. Pereira P, Magnusson B, Theodorsson E, Westgard J, Encarnação P. Measurement uncertainty as a tool for evaluating the “grey-zone” to reduce the false negatives in immunochemical screening of blood donors for infectious diseases. Accred Qual Assur, First Online: November 24, 2015.

20. Ellison SLR, Williams A. Use of uncertainty information in compliance assessment. Eurachem/CITAC, 2007.

21. Pereira P, Westgard J, Encarnação P, Seghatchian J. Analytical model for calculating indeterminate results interval of screening tests, the effect on seroconversion window period: a brief evaluation of theimpact of uncertain results on the blood establishment budget. Transfus Apher Sci 2014, 51(2):126-131.

22. Le T, Bhushan V, Vasan N. First aid for the USMLE step 1. 20th ed. New York (NY): McGraw-Hill Medical, 2009.

23. International Organization for Standardization. ISO 31000 Risk management - Principles and guidelines. Geneva: ISO, 2009.

24. Crofts N, Maskill W, Gust I. Evaluation of enzyme-linked immunosorbent assays: a method of data analysis. J Virol Methods 1988, 22(1):51-59.

25. Pereira P, Westgard J, Encarnação P, Seghatchian J. Evaluation of the measurement uncertainty in screening immunoassays in blood establishments: Computation of diagnostic accuracy models. Transfus Apher Sci 2015, 52(1):35-41.

26. Pereira P, Westgard J, Encarnação P, Seghatchian J, Sousa G. Scheme for the selection of measurement uncertainty models in blood establishments' screening immunoassays. Transfus Apher Sci 2015, 52(1):42-47.

This page intentionally left blank

Chapter 7 - General Conclusions and Suggestions for Further Work

This page intentionally left blank

7.1. General conclusions

The work reported in this thesis shows the relevance of evaluating the measurement uncertainty in screening immunoassays. It allows classifying the results as negative, indeterminate and positive. As with positive results, blood donations with indeterminate results should be rejected. To evaluate the measurement uncertainty at the "cutoff", two intralaboratorial and one EQA/PT novel GUM models were recommended. Despite TAE remains to be used in manufacturers and medical laboratories, it does not comply with the current metrology concepts and therefore it is not an alternative to the GUM approach. The definition of a "rejection zone" including samples with indeterminate results lead to a revision of the definition of the window period, to become the time interval between first infection and when the test result cannot reliably rule out the possibility of infection (due to indeterminate results). This redefinition allows for a shorter window period without increasing post-transfusion risk. It was demonstrated that, when compared to the absolute values, the 95 % CI in diagnostic accuracy estimates provide a more realistic view of risk probabilities, being analogous to the expanded uncertainty concept. Delta-value was presented as a primary measurement of the impact of uncertain results in blood establishments' budget. Finally, an original scheme for the selection of models to evaluate the measurement uncertainty in blood establishments' screening immunoassays was presented.

7.2. Suggestions for Further Work

The results of this research raise a number of issues that deserve further attention:

- In cells and tissues establishments screening of transmissible agents is also a critical step to ensure post-transplant safety. A discussion on suitable methods to evaluate measurement uncertainty and diagnostic uncertainty associated to the used screening immunoassays, like the one that was done in this thesis, could add value to current practice.
- The discussion could be extended to point of care testing (POCT) and rapid tests where measurement uncertainty is not systematically determined and evaluated.

- As with screening immunoassays, measurement uncertainty is not currently evaluated in nucleic acid tests; models to estimate measurement uncertainty in these tests should be discussed and its impact on the interpretation of the tests results evaluated.
- General medical laboratories guidelines do not address models to the determination of measurement uncertainty or present diagnostic accuracy models as tools to the determination of diagnostic uncertainty. Such models could contribute to the accuracy of clinical decisions.
- In other technical contexts of blood establishments, such as in the statistical control of the blood components' production, uncertainty estimation models are of paramount importance to risk assessment.
- Internal quality control schemes of screening immunoassays currently in use do not consider the "rejection zone" when determining the re-evaluation of a sample with results close or equal to the cutoff. New internal quality control schemes that take into consideration the measurement uncertainty at the "cutoff" should be devised.
- Measurement uncertainty could also be reviewed as an analytical hierarchy process approach tool to decision making in blood establishments.